



Learning
for Living

SCHOOL SAMPLING PREPARATION MANUAL

PISA 2006 MAIN STUDY

Version One: April 2005

Project Consortium:

Australian Council for Educational Research (ACER)

Netherlands National Institute for Educational Measurement (CITO)

Educational Testing Service (ETS, USA)

National Institute for Educational Policy Research (NIER, Japan)

Westat (USA)

TABLE OF CONTENTS

1. INTRODUCTION	3
1.1 SAMPLING IN PISA.....	3
1.2 RELATIONSHIP OF FIELD TRIAL SAMPLING PROCEDURES TO MAIN STUDY SAMPLING PROCEDURES	4
1.3 THE RESPONSIBILITIES OF NATIONAL PROJECT MANAGERS AND THE CONSORTIUM IN SAMPLING SCHOOLS.....	5
1.4 A NOTE ABOUT VERSIONS.....	6
1.5 CHANGES FROM PISA 2003	7
1.6 A NOTE ABOUT THE SAMPLING FORMS IN THIS MANUAL.....	7
1.7 THE SCHEDULE FOR SUBMITTING SAMPLING FORMS	7
2. OVERVIEW OF SAMPLE DESIGN	9
2.1 SURVEY GOALS.....	9
2.2 SAMPLING GOALS AND EXPECTATIONS.....	9
2.3 SAMPLE DESIGN OVERVIEW.....	9
3. STEP 1–TARGET POPULATION DEFINITION.....	11
3.1 INTERNATIONAL DESIRED TARGET POPULATION	11
3.1.1. <i>Using table 1 to determine your population definition:</i>	13
3.2 NATIONAL DESIRED TARGET POPULATION	15
3.3 NATIONAL DEFINED TARGET POPULATION	17
3.3.1 <i>Exclusions</i>	17
3.3.2 <i>School-Level Exclusions</i>	17
3.3.3 <i>Within-School Exclusions</i>	18
4. STEP 2 – NATIONAL SAMPLING PLAN.....	23
4.1 INTRODUCTION.....	23
4.2 SAMPLE DESIGN FRAMEWORK	23
4.3 COUNTRY-SPECIFIC ISSUES IN SAMPLE DESIGN	24
4.3.1 <i>Coefficient of Intra-class Correlation</i>	24
4.3.2 <i>Samples of Students in a Particular Grade</i>	25
4.3.3 NATIONAL OPTIONS	27
4.4 PARTICIPATION RATES	27
4.4.1 <i>Schools</i>	27
4.4.2 <i>Students</i>	29
4.4.3 <i>Reporting Participation Rates</i>	30
5. STEP 3 – THE SCHOOL SAMPLING FRAME.....	31
5.1 INTRODUCTION.....	31
5.2 SCHOOL-LEVEL SAMPLING FRAME	31
5.3 AREA-LEVEL SAMPLING FRAME	32
5.3.1 <i>Area-Level</i>	33
5.3.2 <i>School Level</i>	34
5.4 OTHER TYPES OF SAMPLING FRAMES	34
5.5 EXCLUDING SCHOOLS FROM THE SAMPLING FRAME	35
5.6 STRATIFICATION	36
5.6.1 <i>Overview</i>	36
5.6.2 <i>Explicit Stratification</i>	37
5.6.3 <i>Implicit Stratification</i>	37
5.6.4 <i>Stratification Guidelines</i>	38
5.7 TREATMENT OF SMALL SCHOOLS.....	38
5.7.1 <i>Stratification of Small Schools</i>	39
5.7.2 <i>Sample Sizes for the Small School Strata</i>	40
5.8 ASSIGNING A MEASURE OF SIZE (MOS) TO EACH SCHOOL	42
5.9 CERTAINTY SCHOOL STRATUM	43
5.10 STRATA NUMBERING	43
5.11 SAMPLING FRAME STATISTICS	44

STEP 4 – THE SAMPLE OF SCHOOLS	47
6.1 THE SCHOOL SAMPLE	47
6.2 REPLACEMENT SCHOOLS.....	47
6.3 SCHOOL ID’S.....	48
6.4 THE SCHOOL TRACKING FORM	49
6.4.1 <i>Sampled Schools</i>	49
6.4.2 <i>Replacement Schools</i>	51
CHECKLIST OF SCHOOL SAMPLING ACTIVITIES	53
APPENDIX A EXAMPLES OF COMPLETED FORMS	54
APPENDIX B SAMPLING FORM TEMPLATES.....	76
APPENDIX C SCHOOL SAMPLING SELECTION METHODOLOGY	89
C.1 OVERVIEW	89
C.2 SCHOOL SAMPLE ALLOCATION OVER EXPLICIT STRATA	89
C.3 SORTING THE SAMPLING FRAME	90
C.4 DETERMINING WHICH SCHOOLS TO SAMPLE	91
C.4.1 <i>Calculating the Sampling Interval</i>	92
C.4.2 <i>Generating a Random Start</i>	92
C.4.3 <i>Calculating the Selection Numbers</i>	92
C.5 IDENTIFYING THE SAMPLED SCHOOLS.....	94
C.6 IDENTIFYING REPLACEMENT SCHOOLS	95
C.7 ASSIGNING SCHOOL ID’S	96
APPENDIX D OVERLAP CONTROL – PISA AND TIMSS	98
D.1 OVERVIEW	98
D.2 THE OVERLAP CONTROL METHOD.....	98

1. INTRODUCTION

1.1 Sampling in PISA

1. The **Target Population** for PISA is students between 15 years and 3 (completed) months and 16 years and 2 (completed) months at the beginning of the testing window, attending educational institutions located within the country, and in grade 7 or higher. **PISA-Eligible Students** are students who are in the PISA Target Population. To provide valid estimates of student achievement and characteristics, the sample of students must be selected using established and professionally recognised principles of scientific sampling, in a way that represents the full population of PISA-eligible students. Thus the use of appropriate sampling procedures to identify the students who will participate in PISA is very important.
2. Since PISA is a survey of students, and as the sampled students must be given an assessment under standard conditions, in every PISA country it is clear that the sample of students should be selected by first selecting a sample of schools, and then selecting a sample of students within those schools. Thus the need for careful, rigorous, standardised, and documented sampling applies both to the selection of schools, and students within schools.
3. The successful conduct of a sampling plan for PISA can be viewed as having six key components:
 1. The definition of the student population to be surveyed.
 2. The development of a list of schools (more broadly, educational institutions) in which eligible students are, or most likely are, enrolled.
 3. The selection of the sample of schools from this list.
 4. The development of a list of eligible students within each sampled school.
 5. The selection of the sample of students from within each sampled school.
 6. The documentation of the sampling process, and the calculation of indicators of the effectiveness of the sampling and survey operations.
4. This manual describes procedures for achieving points 1, and 2, and, to some extent, 6. Point 3 will be addressed by ACER and Westat. Note that although ACER will select the school samples, all enquiries about sampling must be directed to Westat. Points 4 and 5 are discussed in detail in the PISA Data Management Manual, the PISA National Project Managers Manual, and in the documentation for the KeyQuest software, used for student sampling and data entry, supplied by the Consortium.

-
5. As well as PISA-eligible students, some countries will select a grade-based sample as an international or national option, and this manual provides some discussion about these options. To distinguish between the sampling of PISA-eligible students and the grade-based sampling options, the term **grade-eligible** is sometimes used in this manual. Note that some students may be both PISA-eligible and grade-eligible. However, grade-eligible students are not necessarily PISA-eligible. For example, if the grade-based option relates to students in grade 10, all grade 10 students are grade-eligible, but only those grade 10 students that fall within the ages defined in the PISA Target Population are PISA-eligible.

1.2 Relationship of Field Trial Sampling Procedures to Main Study Sampling Procedures

6. It is very important to note the similarities and differences between the sampling procedures for the field trial and the main study, because while some aspects are very similar between the two, others are very different. Simply following the field trial procedures on a larger scale for the main study will, in most cases, result in a main study sample that is completely unacceptable.
7. The reasons for the differences are that the two components have very different purposes. The field trial had two main purposes:
 1. To collect data to ensure that the instruments developed for the main study contain test and questionnaire items that are sound in all countries and that they are properly translated; and
 2. To test the operational procedures for sampling students and conducting assessments within schools.
8. The main study has the purpose of obtaining a data file that will permit analyses that give valid statistical inferences about the student population, and the characteristics of the schools that they are in. This purpose imposes more rigorous requirements on the sampling procedures than the field trial.
9. The definition of the student population must be much more rigorously applied for the main study than was the case in the field trial.
10. The development of a list of schools, and sampling from this list, requires much more care and preparation for the main study than it did for the field trial, in addition to requiring a larger sample of schools.
11. The procedures for listing and sampling students within schools will be very similar between the field trial and main study. The only changes will be enhancements that are developed in

the procedures as a result of the field trial, or the introduction of some national requirements that did not exist for the field trial.

12. The documentation of the sampling process and the calculation of indicators of quality will be much more extensive for the main study than for the field trial. However, the documentation should largely be a by-product of the frame preparation and sampling procedures. The calculation of indicators will be the responsibility of the PISA Consortium. Thus these aspects should not represent much additional work for National Project Managers (NPMs) and their colleagues on national PISA teams.

1.3 The Responsibilities of National Project Managers and the Consortium in Sampling Schools

13. NPMs are responsible for the following tasks:

1. Establishing the age definition, based on birth date, and the time of testing, according to PISA rules, so as to ensure that the correct student population is surveyed, and to ensure that the assessment can be completed in the designated time period.
2. Determining the geographic, school, and student level exclusions that will apply in their country in PISA, ensuring that these exclusions are kept to a minimum, and documenting the nature and quantity of these exclusions.
3. Obtaining, and where necessary enhancing, a list of schools and other educational institutions in the country that will contain the population of enrolled students that are to be covered by PISA. This list is referred to as the **sampling frame**.
4. Identifying suitable stratification variables for the schools on the sampling frame, and ensuring that these variables are present and essentially correct for all schools.
5. Determining the sample sizes of schools and students that are needed, in conjunction with Westat and ACER, to satisfy international PISA requirements and any additional national requirements.
6. Sending the sampling frame and sample size requirements to Westat and ACER, who will select the sample.
7. Maintaining accurate records as to which sampled schools participate in PISA, the reason for non-participation for each school that does not participate, and the use of replacement schools in the sample.

14. The PISA Consortium, and Westat and ACER in particular, are responsible for the following tasks:

1. Checking that each country has identified an appropriate age definition and time of testing.

-
2. Checking that the exclusions in each country are clearly defined, necessary, and minimal.
 3. Assisting each country in determining the sample size and stratification that will meet both PISA and national objectives.
 4. Selecting the school sample and returning it to the NPM, with documentation.
 5. Developing school and student sampling and replicate weights to permit valid inferences to be made from the sample to the population.
 6. Developing measures of school and student response and exclusion, and coverage of the target population, to assist in evaluating the comparability and quality of the achievement and other data from each country.
15. This sampling manual provides the mechanism by which the Consortium informs NPMs what they need to do to carry out their responsibilities. The sampling forms that are included with the manual provide a mechanism (or at least a template) for NPMs to forward information to Westat and ACER so that we can carry out our responsibilities for monitoring and quality assurance, and sample selection and weighting.
16. For some aspects the manual may document the reason for certain procedures and requirements. But this is not the primary purpose of the manual, and NPMs may need to consult other PISA documents (such as minutes of various meetings) if they wish to understand in detail the purpose of a particular requirement. The prime purpose of the manual is to provide instructions and checklists for completing the necessary tasks without mistakes.
17. One of the responsibilities of Westat is to assist countries in determining a suitable sample design. The only way we could serve this role just through a manual is if a country has no extra national requirements for PISA, and no special circumstances in the student and school population. Since this is rarely the case, NPMs need to contact Westat for individual consultation to ensure that they develop a valid design that meets all the objectives in a cost effective way.

1.4 A Note About Versions

18. Before the PISA main study is completed, there may be two or more versions of this manual produced. The expected difference between these versions is that later versions will include additional material to help inform NPMs about the sampling-related activities of the Consortium and to deal with special topics of interest to only some NPMs.

-
19. In cases where there is a correction or other important change, from one version to the next, each NPM will be explicitly notified about this. We do not expect that NPMs should have to search through each version to find out if there are important differences that affect them.
20. However, please be sure to replace the manual with a new version when it arrives, and **be sure that you are using the latest version of the manual at all times.**

1.5 Changes from PISA 2003

21. Most aspects of this manual are similar to those of the PISA 2000 and 2003 manuals. Where a change of consequence has been made from the 2003 manual, this is noted in the margin as “New for PISA 2006”.

1.6 A Note About the Sampling Forms in this Manual

22. This sampling manual provides paper versions of the sampling forms to record all of the required sampling information for PISA. As will be explained when discussing each form, some of these forms need to be prepared by the NPM and sent to the Consortium for review, while others will be first prepared by the Consortium and sent to the NPM for review.
23. Following agreement, each sampling form will be located on the national profile webpage of the PISA website. Refer to the National Project Managers manual for details of the PISA website.
24. In the preparation of sampling forms to be sent to the Consortium, NPMs need not send versions exactly as they appear in this manual. They may generate their own versions as long as the files include all of the information shown on the sampling forms included in this manual. Similarly, the versions that the Consortium prepares and sends to NPMs may not appear exactly the same as they appear in this manual. However they will contain the same information.
25. The data from these forms will be used in assessing the quality of the PISA data, and therefore must be as accurate as possible.

1.7 The Schedule for Submitting Sampling Forms

26. Following the completion of discussion about the field trial sampling forms, a proposed schedule for submitting the main study sampling forms, linked to the proposed main study

New for
PISA
2006

testing dates, was negotiated with each national centre. The schedule that was negotiated appears on the national centre information page on the PISA website.

27. If any changes are required to this schedule, please inform the Consortium as soon as possible to assist with our planning.

2. OVERVIEW OF SAMPLE DESIGN

2.1 Survey Goals

28. PISA will investigate the knowledge, skills, and competencies of PISA-eligible students in the domains of reading, mathematics, and science. In addition, a representative of each sampled school will also supply information about the school itself and the practices implemented within. Countries will obtain data that will be comparable with all other countries participating in PISA. In addition, national options may be included in the study, provided that they do not interfere with the sampling of the PISA population or the administration of the PISA instrument. Such national options must be approved by the Consortium prior to implementation.

2.2 Sampling Goals and Expectations

29. Survey estimates will be required for students and schools. To obtain reliable estimates for each of these analytical units, quality standards must be maintained with respect to:
1. Accuracy and precision: A minimum sample size of 4,500 assessed students must be selected from a minimum of 150 schools. Further, if a minimum sample size of students is obtained (4,500), the sample must not depart significantly from a self-weighting design. A self-weighting design is one in which each sample unit represents an equal proportion of the total population, when unbiased estimates are made from the sample. This means that if a country wishes to over-sample certain population subgroups, the total sample size will need to be somewhat larger than 4,500 assessed students and 150 participating schools.
 2. Response rates: A response rate of 85 percent is required for initially selected schools as well as a rate of 80 percent of selected students. If the initial school response rate is below 85 percent, an acceptable school response rate may still be achieved through the use of replacement schools (see Section 4.4 for details).
 3. Sampling methodology: National sampling plans must be well-documented and based on sound scientific sampling methods.

30. Refer to the PISA Technical Standards 1.1 – 1.12 and 15.1 for the complete text of all sampling related standards.

2.3 Sample Design Overview

31. The sample design proposed for the PISA assessment is generally referred to as a two-stage stratified sample. The first-stage sampling units consist of individual schools having PISA-eligible students. Schools are sampled systematically with probabilities proportional to a

measure of size (PPS), with the measure of size being a function of the estimated number of PISA-eligible students enrolled. The comprehensive national list of all eligible schools is called the school sampling frame. Prior to sampling, schools in the sampling frame can be assigned to a predetermined number of implicitly and/or explicitly defined strata. A minimum of 150 schools will be selected in each country, with the requirements of national options often requiring a somewhat larger sample. Note that the minimum of 150 schools is selected with the expectation that there will be at least 150 **participating** schools, once field exclusions, ineligibility and nonresponse are accounted for. As the schools are sampled, replacement schools are simultaneously identified, should they be needed to replace nonparticipating sampled schools.

32. The second-stage sampling units are students within sampled schools. Once schools are selected to be in the sample, a list of each sampled school's PISA-eligible students must be prepared. A within-school sample size (referred to as the 'target cluster size') is nominated by each PISA participant, and this number of eligible students is sampled with equal probability from each of the schools sampled for PISA¹, assuming there is at least this number of eligible students at the school. In schools where there are fewer eligible students than the target cluster size, all of these students are sampled. The target cluster size is usually 35 students. In total, a minimum sample size of 4 500 assessed students is to be **achieved**. Note that 30 students (150 schools x 30 students = 4500 students) has been inflated to 35 to account for an anticipated combined nonresponse and exclusion rate of 15% among students within schools. It is possible for countries to negotiate a different target cluster size as a national option, but if it is reduced then the sample size of schools must correspondingly be increased beyond 150, so as to ensure that at least 4 500 students in total were sampled and assessed. The target cluster size has to be at least 20 so as to ensure adequate accuracy in estimating variance components within and between schools—an important analytical objective of PISA.

¹ For different explicit strata, it is possible to vary the target cluster size.

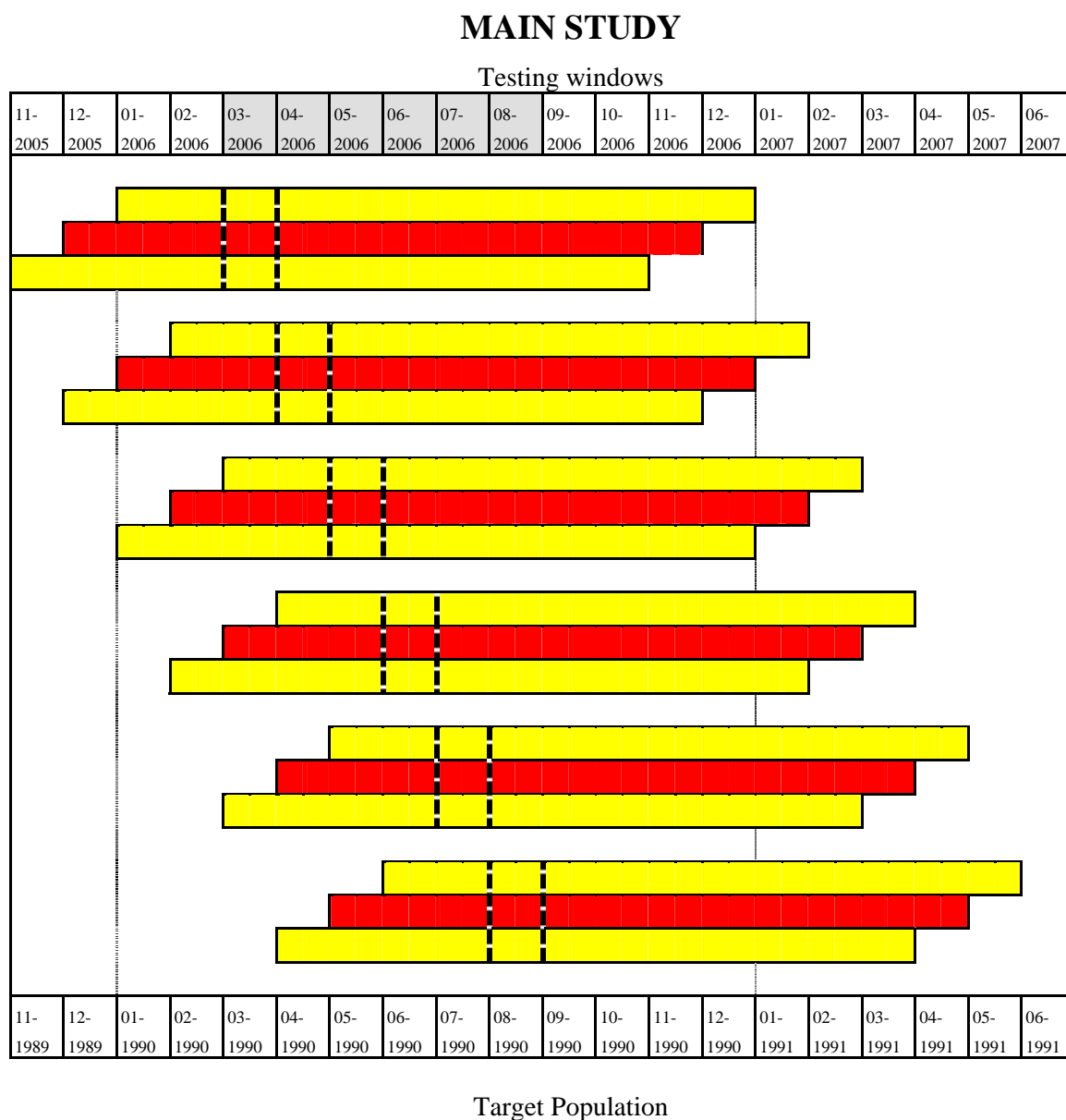
3. STEP 1–TARGET POPULATION DEFINITION

3.1 International Desired Target Population

33. The international desired target population is intended to provide full coverage of all PISA-eligible students in a country's education system. This means that countries are to include PISA-eligible students who attend on only a part-time basis, are in vocational training types of programmes, or any other related type of educational programme, or who are in foreign schools within the country, even if they are not included in other international or national studies. The desired target population does not include residents who attend school in a foreign country.
34. The definition of the population of PISA-eligible students depends in part upon the time of testing within the country. The time of testing depends upon both the period of the academic year within the country, and the convenience of the NPM, taking into account the circumstances within the country that dictate a suitable testing time.
35. The assessment must be conducted during a period not to exceed 42 consecutive calendar days between March 1, 2006 and August 31, 2006 unless otherwise agreed. The assessment should not take place during the first six weeks of the academic year. Thus for many countries in the northern hemisphere, where the academic year begins in September and ends in June, the assessment must be conducted some time between March 1 and the end of the academic year. For countries in the southern hemisphere, that typically have an academic year extending from early February until December, the assessment should be conducted at some time between mid-May and the end of August.
36. As noted, the population of PISA-eligible students is defined by the time of testing, within limits. The standard definition is that the population consists of students attending educational institutions located within the country, in grades 7 and higher, who are aged from 15 years and 3 (completed) months to 16 years and 2 (completed) months at the beginning of the testing period. In the discussions that follow, "attending educational institutions located within the country, in grades 7 and higher" is assumed when talking of the eligible population. This means that if the assessment is to be conducted throughout the month of April 2006, for example, the eligible population is defined as students born during 1990. If the testing is to take place in July 2006, the population is students born between April 1990 and March 1991 inclusive.
37. Variation of up to one month in this age definition is permitted. In particular, if the testing period is any 6-week period between March 1, 2006 and May 31, 2006, the population may

be defined as students born in 1990. (See Table 1.) If no local factors dictate to the contrary, countries are encouraged to test within this time period and to use this population definition.

TABLE 1. SELECTED TESTING WINDOWS AND RELATED AGE DEFINITIONS.



Note:

The red/dark bar shows the standard age definition for each framed test period.

Yellow/light bars show two optional age definitions for each framed test period

3.1.1. Using table 1 to determine your population definition:

Step 1: Look across the top of the table for the first month of your test period.

Step 2: Once found, look downward to find the set of dashed vertical lines that line up with the chosen month. The dashed lines cover three horizontal bars. These bars depict three options for the set of acceptable population defining birth dates.

Step 3: To obtain the birth beginning and ending dates for each of the three options that will correspond to your month of testing, follow the end points of each of the three horizontal bars to the bottom of the table. Record the birth dates.

Step 4: If your testing period involves a second month, repeat steps to 3 for this second month.

Step 5: Then find the birth dates that are in common between the sets of dates obtained in steps 3 and 4. There will be two options available.

Step 6: If your testing period involves a third month the same procedure should be followed for the third month. There will only be one acceptable set of birth dates in this case.

38. Examples of acceptable testing times and age definitions are:

- Testing from March 16 to April 15, students born 12/89 to 11/90
- Testing in April, students born 2/90 to 1/91
- Testing from May 16 to June 15, students born 3/90 to 2/91
- Testing in July, students born 4/90 to 3/91.

39. Examples of testing time and age definition combinations that are NOT acceptable are:

- Testing in April 2006, students born 11/89 to 10/90 (the students are too old).
- Testing in April 2006, students born 3/89 to 2/90 (the students are too young).
- Testing from May 16 to June 15, students born in 1990 (the students are too old).
- Testing in August, students born 7/90 to 6/91 (the students are too young).

40. Although already defined in the planning for the field trial, NPMs must re-confirm their time of testing and age definition with the PISA Consortium, and Westat, in particular. Once the time of the assessment and age definition have been approved, NPMs must be alert to ensure that possible drift in the assessment period does not lead to an unapproved population

definition. For example, an NPM might propose to test during the month of May, students born during 1990. This would be approved. But if in fact the testing period slips to become May to mid June, the population must be changed to students born February 1990 to January 1991 (or March 1990 to February 1991 if preferred), as it is not acceptable to have testing in June with an age definition of students born in 1990. Any changes in the planned time of assessment should be approved by Westat. Note that even though testing is allowed to be conducted over a six-week period, Table 1 shows the testing windows as one-month periods. If the testing covers parts of two months as opposed to just one month, the example just discussed illustrates how to choose the appropriate age definition. One way to minimise the chance of assessment period drift leading to an unapproved population definition is to choose the age definition that will provide flexibility for moderate drift in the test period. Table 1 illustrates how this may be done for an originally scheduled assessment period of April 2006 and an age definition of all students born in 1990 -- the dark bar for the April 2006 testing window. If the testing period shifts to March 2006, the age definition is still valid and is graphically represented by the light-shaded bar at the top of the March 2006 testing window. If the April 2006 assessment period shifts to May 2006, the age definition of all students born in 1990 is also still valid. Graphically, this is represented by the light-shaded bar at the bottom of the May 2006 testing window. Notice, too, that by scheduling a March 2006 test period originally, the age definition of all students born during 1990 will remain an approved population definition even if the testing window is moved to May 2006, and thus provides two months of scheduling flexibility for unexpected delays. This type of approach is recommended.

- 41. Time of testing and age definition must be recorded on Sampling Form 1 and submitted to Westat who will forward the form to ACER. NPMs must submit Sampling Form 1 according to the schedule agreed upon with the consortium at the time of field trial planning. This form should be submitted at least three months before the sample is to be selected.**

Sampling Form 1

NPMs must describe their planned testing period, and definition of PISA-eligible students. On this form NPMs should also indicate any additional grade-eligible students, included as an international or national option (e.g., students in grade 10).

Sampling Form 1- Example

In the following discussion, we show a fictitious example for PISA 2006 of a planned testing period and the subsequent age definition of eligible students using the country of Atlanta.

Due to the academic year in Atlanta beginning in September, the NPM found testing between 15 April 2006 and 15 May 2006 to be the most feasible. This testing date met the PISA criteria that testing must be after the first six weeks of the academic year and within a maximum 6-week testing period. Using the start and end dates of the testing period, the NPM defined the PISA-eligible main study students to be assessed as those born in 1990. As a national option, Atlanta planned to also assess students in grade 12 who were born in 1988, and reported this in question 4. The NPM completed the next three items pertaining to the TIMSS Population 2 study for 2007, indicating that they intended to participate in the study and that the same institution would administer both PISA and TIMSS. The NPM stated that there was no concern if some schools participated in both PISA and TIMSS.

The date this version of Sampling Form 1 was completed was entered on the line below the NPM's name.

See the example of Sampling Form 1 in Appendix A.

Check Point 1

NPMs must submit Sampling Form 1 at least three months before the sample is selected, and according to the agreed upon schedule.

3.2 National Desired Target Population

42. Every NPM must define and describe their country's national desired target population. The national desired target population definition must address the requirements of the international desired target population. It should provide the most exhaustive national coverage of PISA-eligible students as is feasible. Any hardships in accomplishing this must be specified and approved as such, in advance.

Please
Note

43. **The PISA Governing Board (PGB) (formerly known as the Board of Participating Countries (BPC)) and the PISA Consortium strongly encourages all NPMs to provide complete national coverage in their national desired target population.** In cases where the national desired target population deviates from full national coverage of all eligible students, these deviations must be described and enrolment data provided to measure the extent of the reduced coverage. Such situations may occur if, when defining their national desired target population, NPMs find it necessary to reduce their national coverage by removing, for instance, a small, remote geographical region, or a language group, possibly due to political, organizational or operational reasons. Areas deemed by the PGB to be part of

a country (for the purpose of PISA) but which are not included for sampling will be designated as noncovered areas, and must be documented.

44. If such situations occur, the national desired target population will differ from the international desired target population. If a significant proportion of students are excluded from the full national population of PISA-eligible students, this will mean that survey results will not be deemed representative of the entire national school system.

Please
Note

45. **The national desired target population must be described on Sampling Form 2 and submitted to Westat who will forward the form to ACER.** By defining the national desired target population on Sampling Form 2, and describing any limitations in national coverage, this information will be properly documented and can be formally and adequately presented in the international reports.

Sampling Form 2

NPMs must describe their national desired target population by specifying enrolment statistics for PISA-eligible students. **Note that although PISA-eligible students include mostly 15-year-olds as well as some students who are 16-year-old, the best information that most countries will have available at the time of completion of sampling form 2 is population and enrolment statistics about 15-year-olds. Thus, this is the terminology used on sampling form 2.** NPMs are requested to describe the extent of national coverage in their national desired target population. Any populations excluded from the full national population of 15-year-olds must be documented on this form.

Sampling Form 2- Example

The following example is based on a fictitious Atlanta plan for PISA for the year 2006 main study. Using information from the *2004 Database of the Statistical Yearbook of Atlanta (school year 2004/2005)*, the NPM found the total national population of 15-year-olds to be 231645. Of this total, 226658 15-year-olds in grades 7 and higher were noted as enrolled in educational institutions.

The NPM reported in question 2.2 that no 15-year-olds were enrolled in grades 6 and below. Of those students in grades 7 and above, 0 enrolment was to be excluded. Therefore, total enrolment of Atlanta's desired target population was all 226658 15-year-old students. Finally, the NPM noted the source used to define the desired target population.

The date this version of Sampling Form 2 was completed was entered on the line below the NPM's name.

New for PISA 2006, the webpage containing the required information was provided for reference.

See the example of Sampling Form 2 in Appendix A.

3.3 National Defined Target Population

46. Using their national desired target population as a general framework, NPMs must then describe their national defined target population. The national defined target population is essentially the school sampling frame. Differences between the national desired and defined target populations will generally arise out of practical considerations and should be kept to a minimum.

3.3.1 Exclusions

47. The national defined target population may represent a subset of the national desired target population. All schools and students from the national desired target population excluded from the national defined target population are referred to as the excluded population.

48. Usually, practical reasons are invoked for excluding schools and students, such as increased survey costs, increased complexity in the sample design and/or difficult test conditions. Exclusions can occur at the school level, i.e., entire schools are excluded, or within schools, i.e., specific students within sampled schools are excluded. All such exclusions, at both the school level, and the within-school level, must be described and quantified in Sampling Form 3.

49. In PISA, exclusions from the national desired target population are to be kept to a minimum. Unlike some other international education studies, difficulties raised by the existence of small schools, (i.e., those with few PISA-eligible students), or other schools where it is difficult or costly to conduct assessments, are mainly to be addressed by modifying the sample design to reduce the number of such schools selected, rather than by exclusion. Similarly, the exclusion of special education students and students with insufficient assessment language experience² is to be kept to a minimum.

3.3.2 School-Level Exclusions

50. NPMs may wish to exclude a very small fraction of eligible schools for the following reasons:

1. the school is geographically inaccessible (but not part of a region that is omitted from the National Desired Target Population); or
2. administration of the PISA assessment within the school would not be feasible.

² Refer to definition 15 of the PISA Technical Standards

51. The percentage of the target PISA-eligible population enrolled in such excluded schools, in total, should be less than 0.5 percent of the total population of PISA-eligible students. Note that extremely small schools (i.e. those with exactly only one or two PISA-eligible students) may be excluded in the field after initial contact IF the number of students in these schools on the frame accounts for no more than 0.5 percent of the total population of PISA-eligible students, AND there is still room left in the 0.5 percent exclusion limit. These schools should not be excluded outright from the frame. Instead, this method is adopted to guard against erroneous size measures leading to more than expected excluded students.

52. In addition, a school may be excluded if it provides instruction only to students in the excluded categories defined under “within-school exclusions” (described in the next section), such as schools for the blind. The percentage of PISA-eligible students enrolled in such schools should be less than 2 percent.

Please
Note

53. **The magnitude, nature, and justification for school-level exclusions must be documented on Sampling Form 3 and submitted to Westat who will forward the form to ACER.** One of the quality indicators produced by PISA will be based on school-level exclusions.

54. Note: The population of PISA schools is limited to schools that could contain PISA-eligible students. Schools that do not contain any such students, such as lower primary schools in some countries, for example, are therefore not considered part of the excluded school-level population.

3.3.3 Within-School Exclusions

55. Within-school, or student-level exclusions will provide another indicator of the quality of the national PISA samples. Because definitions of within-school exclusions will vary from country to country, NPMs are asked to adapt the following rules so that they are workable in their country and coded according to the PISA international coding scheme.

56. Within schools, all PISA-eligible students should first be listed. Sampled students who are deemed as excluded must be retained, and a variable maintained to briefly describe the reason for exclusion (see PISA National Project Managers Manual). Using this method, the size of the within-school exclusions can be well estimated from the sample data.

57. International within-school exclusion rules are specified as follows:

1. Functionally disabled students – These are students who are permanently physically disabled in such a way that they cannot perform in the PISA testing situation. Functionally disabled students who can respond should be included in the testing.

-
2. Intellectually disabled students – These are students who have a mental or emotional disability and who, in the professional opinion of qualified staff are cognitively delayed such that they cannot perform in the PISA testing situation. This includes students who are emotionally or mentally unable to follow even the general instructions of the test. Students should not be excluded solely because of poor academic performance or normal discipline problems.
 3. Students with insufficient assessment language experience to take the PISA test— Students with insufficient assessment language experience are students who meet all of the following criteria: a) are not native speakers in the assessment language b) have limited proficiency in the assessment language c) have received less than one year of instruction in the assessment language. Only students meeting all three of these criteria can be excluded for this reason.

Please
Note

58. The exact extent of within-school exclusions will not be known until the within-school sampling frames have been returned from the participating schools and sampling weights computed. **Estimates of the extent of within-school exclusions are required from the NPMs to ensure that the national defined target population will cover at least 95 percent of the national desired target population. At the time of frame creation, NPMs are asked to provide their best estimate of the extent of within-school exclusions by type of exclusion and report the estimates on Sampling Form 3.** Only a single within-school exclusion category for type "Other" should be needed within a country, and this exclusion category must be defined. Although it is expected that the three defined categories should cover all types of within-school exclusions, it may be the case that a particular country has an additional category. Use of such a category is described in the NPM manual, but must be limited to special circumstances as defined by the NPM and approved by the Consortium.

59. NPMs should note the distinction to be made between within-school exclusions and non-response. Students unable to perform the achievement tests because of a permanent condition should be excluded. Students with a temporary condition at the time of testing, such as a broken arm, will be treated as non-respondents along with other “absent” or refusing sampled students.

Please
Note

60. The sum of the school-level and within-school exclusions must be limited to 5 percent of the national desired target population. **This value though, should not be treated as a limit below which anything is completely acceptable.**

61. The extent of such exclusions will be quantified to produce one of the quality indicators produced from the study. This indicator will be the proportion of the National Desired Target Population that is covered by the study. The numerator of this proportion will be estimated unbiasedly from the sample (and will thus be subject to sampling error), while the denominator will be the count of the number of persons born between the target population birth dates, within the Desired Target Population, from official demographic statistics for each

country. This proportion will fall below 1.0 in each country when there are youth whose age falls within the PISA population definition, but who do not attend an educational institution, or who do but are in grades 6 or below, excluded schools, and/or excluded students. This provides a measure of the extent of deviations, in terms of reduced coverage, from the national desired target population, combined with the proportion of students not enrolled. This coverage measure will be accompanied by a measure of the size of the difference between the international and national desired populations.

62. Table 2 presents a summary of exclusion limits.

TABLE 2. EXCLUSION LIMITS SUMMARY.

Exclusion Type	Reason	Upper Limit
From the International Desired Target Population	Political, organizational, operational	Not specified – must be discussion with and approval by Consortium. Specify on Sampling Form 2.
From the National Desired Target Population:		
School-Level:	Geographical inaccessibility; Administration of PISA not feasible; <i>(Extremely small size but only in the field and not from the frame)*;</i>	0.5% of National Desired Target Population. Specify on Sampling Form 3.
Within-School Level:	All students of types allowed for within-school exclusions if whole school contains only such students Functionally disabled students; Intellectually disabled students; Students with insufficient assessment language experience	2% of National Desired Target Population. Specify on Sampling Form 3. <i>Estimate</i> expected to be within 2.5% of National Desired Target Population Specify on Sampling Form 3.
TOTAL	-	5% - National Defined Target Population should cover 95% or more of the National Desired Target Population

*Note that extremely small schools (i.e. only 1 or 2 PISA-eligible students) cannot be listed as exclusions in question 2 on sampling form 3 since they cannot be excluded from the school frame. However, if their students fit within the 0.5 percent exclusion limit after all other exclusions under the 0.5 percent exclusion limit are taken into account, and approval has been given, they can be excluded in the field during data collection and will be counted as part of the school-level exclusion estimate.

Sampling Form 3

NPMs must describe their national defined target population by describing all types of school-level and within-school exclusions. The size of these excluded populations must also be reported.

Sampling Form 3 – Example

The following fictitious discussion uses Atlanta as an example. The national desired target population was composed of 226658 15-year-old students. After excluding 145 schools, the coverage of the national desired target population was reduced by 1.1 percent.

In this example, within-school exclusions of all types were expected to be rare, and therefore deemed as negligible. After exclusions, the national defined target population covers 98.9 percent of the national desired target population.

Each type of school-level exclusion is numbered. In this example, school-level exclusion type 1 corresponds to "Special schools for students with special education needs." The example for Sampling Form 12 illustrates how this exclusion type number is to be used for sample participation status coding. (See Section 6.4.1 for details.)

The date this version of Sampling Form 3 was completed was entered on the line below the NPM's name.

New for PISA 2006, the webpage containing the required information was provided for reference.

See the example of Sampling Form 3 in Appendix A.

Check Point 2

NPMs must submit completed Sampling Forms 2 and 3 to Westat, according to their agreed upon schedule, and generally three months prior to the date that they wish to have their school sample selected.

4. STEP 2 – NATIONAL SAMPLING PLAN

4.1 Introduction

63. The development of a national sampling plan is the responsibility of every NPM. National sampling plans will need to be adapted to the particular requirements of national school systems. The staff at Westat can assist NPMs in developing their national sampling plans. The final national sampling plans will also need to be approved by Westat. This approval will be needed before the school samples are selected by ACER.
64. The PGB has set high standards for coverage, response rates (85% at school level and 80% at student level) and the quality implementation of sampling plans as noted in Section 2.2. These high standards are the assurance that the resulting samples will be of the highest quality thus yielding unbiased, accurate and internationally comparable survey estimates.

4.2 Sample Design Framework

65. The general sample design framework adopted for this study is a two-stage stratified sample design. (A preliminary extra first stage is discussed in Section 5.3, for those countries needing it.) National sampling plans must be based on sound and defensible sampling methods. This sampling manual is intended to describe how to implement such methods. The key features are briefly described in the following points.

1. A sampling frame must be developed according to guidelines in Chapter 5.
2. For the first stage of sampling, schools will be stratified, explicitly and/or implicitly. At least 150 schools will be sampled with a sampling method called probability proportional to size (PPS) systematic sampling. Note that it is expected that there will be a minimum of 150 responding schools. Thus, the minimum number of schools selected needs to take expectations about field exclusions, ineligibility, and nonresponse into account. Although most countries will not need this, a description of this sampling method can be found in Appendix C. If explicit stratification is used, the overall sample size will need to be allocated among the explicit strata. *Westat and ACER will perform this function in close consultation with the NPM.* Countries are permitted to sample more than 150 schools, to meet national requirements. In countries where a sizeable proportion of PISA-eligible students are in schools with only a few PISA-eligible students and for countries with highly tracked school systems at this age (where track is not used in stratification), the school sample must be appropriately increased. In the latter case, this is to allow for the considerable school-to-school variation in achievement that can be anticipated. Such an increase is not required, however, in highly tracked school systems at this age, if the different tracks are included in distinct explicit or implicit strata. **For this reason, stratification by school level and/or track is strongly encouraged.** Another issue affecting the number of schools for selection is that of small schools. Small schools

New for
PISA
2006

are a concern because fewer than the desired number of students are sampled from such schools. This leads to a reduction in student sample size. Therefore, small schools may require special treatment in terms of stratification, and perhaps their selection probabilities, and an overall increase in the school sample size. This issue is discussed in more detail in Section 5.7.

Please
Note

3. The second stage of sampling consists of selecting an agreed number (the target cluster size, negotiated on field trial sampling form 2), usually 35, PISA-eligible students from each sampled school. Recall that 35 was the inflated number from 30 per school to compensate for an expected 15% within school nonresponse and exclusion rate. If countries opt to select more than 150 schools (with no other changes in sample design), and have the expectation that all, or nearly all of the selected schools will participate, they will be able to select fewer students per school as long as the total student sample size of about 4500 assessed students is still obtained. If, however, NPMs wish to change the sampling plan so that they also select samples of 10th graders, for example, who are not PISA-eligible, the sample sizes of schools and students, and the sample design itself, will need to be modified so that the PISA international standards are not compromised. Westat will provide assistance for those countries wishing to do this. (See more in Section 4.3.2)

New for
PISA
2006

4. For countries that participated in previous cycles of PISA which had larger than anticipated sampling variances associated with their estimates, or had fewer than 4500 assessed students, recommendations will be made about sample design changes that will help to rectify these problems for PISA 2006. Additionally, although information was supplied on field trial sampling forms, countries will be asked through email exchanges about anticipated school and student response rates, as well as about rates of both school and student ineligibility. Westat can take this information into account when determining school sample sizes to guard against low student yield.

4.3 Country-Specific Issues in Sample Design

4.3.1 Coefficient of Intra-class Correlation

66. The coefficient of intra-class correlation (IC) provides a measure of the tendency of student characteristics to be more homogeneous within schools, or within classrooms, than would be the case if students were assigned randomly to schools or classrooms.
67. Multi-stage sample designs, such as the one for PISA, are affected by the IC. As the value of the IC increases, multi-stage sample designs become less efficient in producing reliable survey estimates, for a fixed sample size of students. In other words, sample sizes must increase as the IC increases to obtain equivalent levels of sampling precision. Depending on the value of the IC in a particular school system, either more schools with fewer students selected within schools, or fewer schools with more students selected within schools, are needed in the sample design to meet the data reliability requirements (i.e. reasonable standard errors for estimates).

-
68. The value of the IC may be approximated from previous surveys that employed similar target populations, similar sample designs, and similar criterion variables. The values for the IC tend to be higher for designs where classrooms are sampled rather than for designs where students are directly sampled within schools. The difference is generally much larger in school systems where the students are streamed, or tracked, within schools.
69. For most countries which will follow the standard PISA two stage stratified cluster design, the IC has already been accounted for in determining a minimum school sample of 150 participating schools and a minimum student sample of about 4500 assessed students. However, these points are relevant for PISA in the situations where countries may wish to oversample schools, and in countries who desire to sample students within grade 10, for example, who are not PISA-eligible.

4.3.2 Samples of Students in a Particular Grade

70. The proposal put forth by the Consortium for PISA 2006 responded to a discussion in the call for tender about an international grade-based sampling option. The grade identified for each country is the one with the greatest proportion of PISA-eligible students. In countries where two grades each have more than 40 percent of these students, the NPM can indicate which of the two grades they wish to include. In such cases, if a country elects to include both grades, the second grade becomes a national option, rather than an international one. In fact, any country can elect to add an additional grade of its own choosing, as a national option.

New for
PISA
2006

71. The option calls for the total school and student sample size of assessed students for the grade-based component to meet the same minimum requirement as for the PISA student assessment. Thus, there must be at least 4500 assessed students from 150 participating schools. If direct student sampling is used, the results from this sample will be of comparable precision to the age-based results, with perhaps some increase in precision as the population is somewhat more homogeneous in some countries.
72. For each country, the increase in the sample size depends upon whether or not direct student sampling is used. In such a case, the increase in sample size then depends on the overlap between PISA-eligible students and the modal grade of PISA-eligible students as well as on whether more students will be selected per school, and/or more schools selected. These options were discussed at previous NPM meetings and two tables were produced indicating the approximate increase in number of schools to be sampled, or students to be sampled per school, for each country. These two tables can be found on the Consortium website under PISA 2006, Meetings_06, Athens March 2004: NPM(0403)8Supp_1.

-
73. With the student-level sample, the sample of grade-eligible students would be sampled directly from a list of all such eligible students within the school. This sample selection could, and most likely would, be fully integrated with the selection of the PISA-eligible sample. Thus, participating schools would be asked to provide a student list containing all PISA-eligible students, plus all other students enrolled in grade 10, for example. A sample of the appropriate size would be selected from this combined student list, where the size would be determined in part by the extent of overlap between the PISA-eligible and grade-eligible populations. The two samples, the PISA-eligible sample and the grade-eligible sample, will be weighted together. For analysis purposes, the students can be separated into PISA students or grade students as desired. Students who are eligible for both PISA and the grade sample can be used in both analyses, and will have the same initial base weights in each.
74. Although direct student sampling is generally preferable, a classroom-level option is another way for the international grade option to be achieved. In the case of classroom sampling, the selection of one or more classrooms from within the selected schools would be appropriate. Thus, the sample selection within a school would consist of two components. First a sample of PISA-eligible students would be selected from a list of PISA-eligible students enrolled in school. Next, a sample of grade 10 (for example) science classes would be selected from a list of grade 10 science classes that covered all grade 10 students in the school. All students from the selected classes would be included in the sample. Some students, who are both PISA-eligible and grade-eligible, would no doubt be selected twice. Such students would not need to be assessed twice, but the data they provide from a single assessment could be used in analyses of both the grade- and age- eligible populations, with different student weights (because of different sampling methods) for each type of analysis.
75. The classroom-based approach will present challenges in many cases, in the preparation of a suitable list of classrooms from which to select the sample. Also the best approach will depend upon whether classroom-based sampling is being used purely for operational convenience, as the easiest way to administer the assessment to a sample of grade 10 students (although this is unlikely to be the case) or whether the aim is to gather information about classrooms of students, either in association with a teacher questionnaire, or just as a means to provide good data for use in variance decomposition and mixed linear models (HLM).
76. Countries interested in having an international grade option must alert Westat to this fact when discussing the school sample.
77. Countries wishing to have a grade sample in a sample of schools that is different than the PISA sample of schools should handle this independently of PISA, although such a national option is still subject to Consortium approval as described below.

4.3.3 National Options

78. All national options must be approved by the consortium, through other forms (such as the field trial IOP forms) not included in this manual, to ensure that they will not be detrimental in any way to the main PISA study. Additionally, if a country will include national options in their main study, they should check with Westat to ensure that their plans do not need to be taken into account in sample design.

4.4 Participation Rates

79. The PISA data quality standards require minimum participation rates for schools, as well as for students. These standards exist to minimise the potential for response biases.

4.4.1 Schools

80. PISA requires a minimum participation rate of 85% of originally sampled schools. However, nonparticipating sampled schools may be substituted with “replacement schools” to meet sample size and response rate requirements. The use of replacement schools does not guarantee that potential biases have been reduced. Therefore, NPMs are encouraged to persuade as many original sampled schools as possible to participate; only a high participation rate will minimize the potential for response bias. The identification and use of replacement schools are described in sections 6.2 and 6.3. Note that raising participation rates, also known as response rates, through the use of replacements improves quality, but that a given response rate achieved through the use of replacements is not as good as that same rate achieved without replacements. That is, a school participation rate of x% before replacement has less potential for biases than the same response rate of x% achieved only after replacement. Therefore, all other factors being equal, acceptability of the country's data in international comparisons will be relative to both the school participation rate of originally sampled schools and the response rate achieved with the use of replacements.

Please
Note

81. Three response rate zones -- acceptable, intermediate and not acceptable -- are defined. “Acceptable” means that the country's data will be included in international comparisons. “Not Acceptable” means that the country's data will be a candidate for not being reported in international comparisons, and will be included only if the NPM provides considerable evidence that nonresponse bias is likely to be minor. Finally, the “Intermediate” zone means that a decision on whether or not to include the country's data in international comparisons must be made while taking into account various other factors. These zones are a function of the response rate before replacement, and the response rate after replacement. There are three

sets of combinations of before and after response rates that are noted as being in the “acceptable” zone.

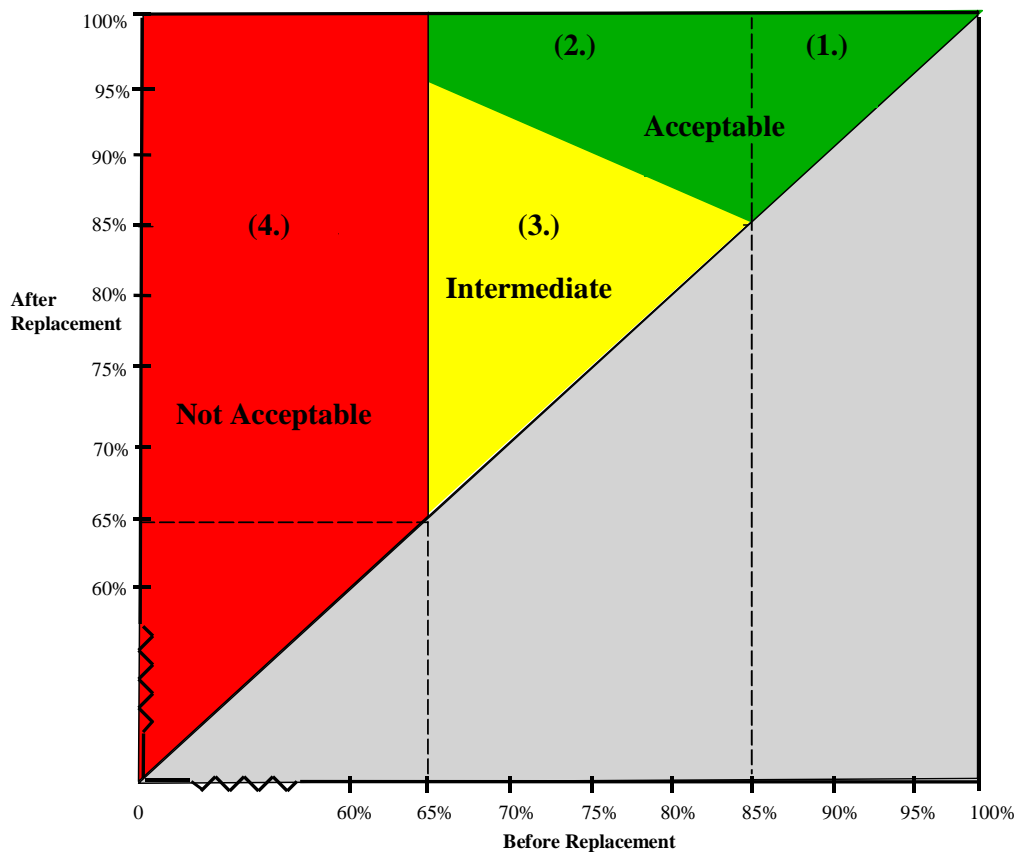
82. Consider some examples.

- A country with a before replacement response rate of 87% would have its data considered “acceptable” according to this criteria, even before any school level replacements were used.
- A country with a before replacement response rate of 60% would have its data “not acceptable” no matter what the after replacement response rate was.
- A country with a before replacement response rate of 70%, and an after replacement response rate of 96% would have its data considered “acceptable” according to this criteria.
- A country with a before replacement response rate of 75% and an after replacement response rate of 91% would have its data considered “acceptable” according to this criteria.
- A country with a before replacement response rate of 70%, and an after replacement response rate of 90% falls into the intermediate zone. Such a set of conditions requires a decision on whether or not to accept the data from the country. Other factors that will be considered in such a situation include student-level response rates, exclusion rates, quality control data, and closeness of the response rates to the acceptable level.
- The response rate zones are shown diagrammatically in the following graph. Note that the “acceptable” region has part of its region bounded by the Before Replacement Rate $> 85\%$ vertical line, another part consisting of the After Replacement Rate $> 95\%$ horizontal line for Before Replacement Rates between 65% and 85%, and that the final part of the acceptable region is defined by the Before Replacement Rate between 65% and 85% and After Replacement Rate $> (255\% - \text{Before Replacement Rate})/2$.
- Note that a school with less than 50% participation among the selected eligible and non-excluded students will not be considered as a participating school. If such a school has less than 25% participation among the selected eligible and non-excluded students, then the students in such schools will not be included in analysis. If such a school has more than 25% and less than 50% participation among the selected eligible students, the students in these schools will be retained for analysis, even

Please
Note

though the school is considered a non-participant for the purposes of monitoring response rates. Note that “selected eligible students” are those in the target population, and therefore participation measures are relative to only those students in the target population. For example, participation will be determined separately for the PISA-eligible portion and the grade-eligible portion of an international grade option sample in the same school. This could result in the school deemed as participating (50% or more of selected grade-eligible students participated) for the grade sample, but as a nonparticipating school for PISA if less than 50% of the selected PISA-eligible non-excluded students participate. Also note that a selected PISA-eligible student recorded in the database as not doing the cognitive part of the PISA assessment will be counted as a nonparticipant for the 50% and 25% participation measures, even if the student completed the background questions.

PISA 2006 School Response Rates



4.4.2 Students

83. PISA also requires a minimum participation rate of 80 percent of students within participating schools (sampled and replacement). Follow-up sessions may be required in schools where

too few students participated in the originally scheduled test sessions. Guidelines for determining when follow-up sessions are appropriate are found in the PISA National Project Managers Manual. The task of determining whether follow-up sessions are required will be delegated to the School Coordinators and Test Administrators. As they oversee the actual test sessions, they will be able to promptly determine whether follow-up sessions are required at each school.

84. Student participation rates are calculated over all participating schools, whether sampled schools or replacement schools, and from the participation of students at the originally scheduled session and any follow-up session that may be required. The student participation rate requirement must be met at the national level, not necessarily for each participating school.

4.4.3 Reporting Participation Rates

85. National participation rates will be presented in the PISA analytical reports. They include:
- Weighted and unweighted school participation rates with and without replacement schools (minimum 85% required without replacement schools); and
 - Weighted and unweighted student participation rates (minimum 80% required).
86. The PISA analytical results for each country will be annotated, based on whether or not the response rate requirements have been adequately met.

5. STEP 3 – THE SCHOOL SAMPLING FRAME

5.1 Introduction

87. This step requires all NPMs to construct a school sampling frame covering their national defined target population. A well-constructed school sampling frame is one that provides complete coverage of the national defined target population without being contaminated by incorrect entries, duplicate entries, incomplete entries, or entries that refer to elements that are not part of the defined target population. Initially, this list should include any school that could have PISA-eligible students, even those that may later be excluded (see Section 3.3.2). The quality of the sampling frame has a direct effect on the survey results, and thus frame construction should be treated with extreme care. PISA will assess the extent of missing or inaccurate data used for stratification (Section 5.6) and for creating the measure of size (Section 5.8), as these will increase sampling error.

88. The construction of a school sampling frame depends to a great extent on the availability of appropriate information about schools and students. The main types of sampling frames that are likely to be used in this study are described below. **NPMs will be asked to describe their choice of a sampling frame on Sampling Form 4 and to submit this form to Westat who will forward the form to ACER.**

Please
Note

5.2 School-Level Sampling Frame

89. School-level sampling frames are based on a comprehensive national list of schools. That being said, how is a “school” defined? Although there is no easy answer to this, we generally wish to sample whole schools as the first stage unit of selection, rather than programmes or tracks within schools, or shifts within schools, so that the meaning of “between school variance” is more comparable across countries. There are exceptions to this, such as when school shifts are actually more like separate schools than part of the same overall school. However, in some countries with school shifts this is not the case and therefore whole schools are used as the primary sampling unit. Similarly, many countries have schools with different tracks/programs but generally we recommend again that the school as a whole should be used as the primary sampling unit. There are some exceptions, such as the schools being split for sampling in previous PISA cycles (trends would be affected if the same practice was not continued), or if there is a good reason for doing so (such as to improve previously poor response rates, differential sampling of certain tracks or programs is desired, etc.). Further

discussion in this section uses the term “primary sampling unit” in place of “school” so that this is all-inclusive.

90. An approximate enrolment (ENR) of PISA-eligible students (i.e. 15-year-olds) is associated with each school listed in the sampling frame. The quality of a sampling frame will, to a large extent, depend on the accuracy of the ENR available. A suitable school ENR value is a critical component of a school sampling frame since school selection probabilities are based on this quantity. The ENR gives a more or less accurate indication of the number of eligible students in each listed school. The best school ENR for PISA would be the number of PISA-eligible students currently enrolled. Current enrolment data, however, is rarely available at the time schools are sampled. If none of the types of ENR listed below are available, or if the available enrolment data are too out of date, schools may have to be selected with equal probabilities, which may in turn require an increased school sample size.

91. Each school entry on the frame should include at minimum:

1. School identification information, such as a unique numerical national ID, and contact information such as name, address, phone number, etc.
2. A suitable measure of the approximate enrolment of PISA-eligible students (ENR). In order of suitability, examples of this are:
 - Student enrolment of 15-year-olds;
 - If 15-year-olds tend to be enrolled in two or more grades, and the proportions of students who are 15 in each grade are approximately known, the 15-year-old enrolment can be estimated by applying these proportions to the corresponding grade level enrolments;
 - The grade enrolment of the modal grade for 15-year-olds; and
 - Total student enrolment, divided by the number of grades in the school.
- Coded information about the school, such as region of country, school type (public or private), urban/rural classification, etc. which may have some effect on the assessment outcomes, and can be used as stratification variables (variables used for dividing the population into mutually exclusive groups so as to improve the precision of sample-based estimates) (See Section 5.6.).

5.3 Area-Level Sampling Frame

92. Area-level sampling frames are used when a comprehensive national list of schools is not available or cannot be constructed without undue burden. As a consequence, area-level sampling frames introduce an additional stage of frame creation and sampling (called the first stage of sampling) before actually sampling schools (the second stage of sampling). This manual provides only an overview of how to implement this additional stage of sampling.

However, the school sampling methods presented in the manual can be adapted for that purpose. Furthermore, Westat will provide support to NPMs that require the use of an area-level sampling frame. NPMs using an area-level sampling frame should submit to Westat, a sampling form 3 for the full national defined population as well as one for the population numbers corresponding to the sampled first-stage units.

Please
Note

93. The use of an area-level sampling frame increases the degree of clustering in the sample. This could be cost effective in countries where distances are great and/or travel costs are high. Area-level sampling frames, however, **usually require a larger sample size** because of this increased clustering. NPMs are encouraged to make every effort to locate a comprehensive national list of schools that could produce a school-level sampling frame.

5.3.1 Area-Level

94. Sampling units on this type of sampling frame are usually geographical areas (GAs), for which reliable information on student enrolment is known. The GAs are usually identifiable geographical units used during the taking of a national census, or administrative units for which basic education statistics are available.
95. The GAs become the primary sampling units (PSUs) at this first sampling stage. The area - level sampling frame then, consists of a list of all eligible PSUs. There needs to be a sufficiently large number of PSUs to permit the random selection of a minimum number of PSUs in the sample. This minimum number will ensure that the sampled PSUs can effectively be considered as representative of all PSUs. The PISA guidelines on this matter require a minimum of 80 PSUs in the area-level sampling frame and a minimum of 40 PSUs sampled from that frame using PPS sampling with the size ideally being the number of 15-year-olds. (Exceptions regarding these minimums must be discussed with Westat.) NPMs will need to discuss the implementation of these guidelines with Westat, based on their specific circumstances. The inability to meet these guidelines will mean that the sampled PSUs can only be considered representative of themselves, rather than representative of all PSUs. The result would be a potentially low national coverage, which would be reported, as such in the analytical reports.
96. Each PSU in the list should include at minimum:
- Unique PSU identification information;
 - Appropriate entries for any suitable stratification variables (see Section 5.6); and
 - A suitable PSU measure of size.

97. Suitable PSU measures of size could be:

- 15-year-old student enrolment;
- Total student enrolment;
- Number of schools; or
- Population size.

5.3.2 School Level

98. The second level of this type of sampling frame consists of a comprehensive list of schools from all sampled PSUs. For each PSU selected from the area-level sampling frame, a comprehensive list of schools is prepared according to the guidelines in Section 5.2.

5.4 Other Types of Sampling Frames

99. In some special cases, NPMs may wish to use a sampling frame that is not related to either type described above. All such cases must be approved by Westat.

Sampling Form 4

NPMs must describe the type of sampling frame they intend to use.

If an area-level sampling frame, or any other type, is to be used, NPMs must provide a description of the information available to construct this frame. For an area-level sampling frame, this information should be a description of PSUs and their total number, the PSU measure of size available and how many PSUs they intend to sample.

NPMs should also describe the school estimate of enrolment they intend to use.

New for PISA 2006, NPMs are asked to describe the units on the school sampling frame (i.e. whole schools, shifts, tracks, programmes, etc.) If the country previously participated in PISA, they are also asked to indicate if the current unit type is the same or not as the units used in previous PISA cycles.

Sampling Form 4 – School-Level Sampling Frame Description Example

Consider a fictitious example using Atlanta. A comprehensive national list of schools was available and therefore a school-level sampling frame was used. The school estimate of enrolment consisted of enrolment of 15-year-olds. Enrolment data from the 2004-2005 school year was available for conducting PISA in the year 2006.

The date this version of the sampling form was completed was entered.

New for PISA 2006, Atlanta indicated that it is whole schools that are listed on the school sampling frame, and that this is the same as what has been done in previous PISA cycles.

This information is reported in the first of the examples of Sampling Form 4 in Appendix A.

5.5 Excluding Schools From the Sampling Frame

- 100.** Based on the National Defined Target Population information reported on Sampling Form 3, some schools may need to be excluded from the school sampling frame. **These schools and their enrolment sizes of 15-year-olds must be identified on Sampling Form 5, along with the reason for their exclusion, and this form must be sent to Westat who will forward the form to ACER.**

Please
Note

- 101.** NPMs should keep in mind that the proportion of eligible students in all excluded schools should not exceed 2.5 percent of the national desired target population. NPMs may have chosen to exclude geographically remote schools. This choice may very well depend on being able to keep exclusions within the 2.5 percent limit. Excluding extremely small schools from the school frame is not permitted (see Sections 3.3.2 and 5.7).

Please
Note

Sampling Form 5

NPMs must list and describe all schools that will be excluded from their school sampling frame. This can be done by simply removing these excluded schools from the school sampling frame and recording them in a separate computer file.

For each excluded school the reason for exclusion and the number of eligible students enrolled (ENR) should be recorded.

Sampling Form 5 - Example

Using Atlanta as an example, the school identifiers of the schools that are to be excluded because they contain only students with special needs, are listed, along with the approximate enrolment (ENR) for each.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

This information is recorded in the example of Sampling Form 5 in Appendix A.

5.6 Stratification

5.6.1 Overview

102. Prior to sampling, schools can be ordered, or stratified, in the sampling frame. Stratification consists of classifying schools into groups according to some stratification variables. Stratification is generally used for the following reasons:

- To improve the efficiency of the sample design, thereby making survey estimates more reliable;
- To apply different sample designs, such as disproportionate sample allocations, to specific groups of schools, such as those in states, provinces, or other regions;
- To ensure that all parts of a population are included in the sample;
- To ensure adequate representation of specific groups of the target population in the sample; and
- To obtain reliable estimates for each stratum, if so required.

103. Examples of stratification variables include, but are not limited to:

- Regions (states, provinces);
- Urbanisation (rural areas, urban areas);
- Socio-economic status (low, medium or high income);
- School types (public, private);
- School size; and
- School programmes (academic, vocational).

5.6.2 Explicit Stratification

104. Explicit stratification consists of dividing the sampling frame (i.e. the list of all schools that could contain PISA-eligible students) into mutually exclusive strata that are each sampled separately. This division of the sampling frame into separate strata will be undertaken by the Consortium using the levels of the explicit stratification variables recorded by the NPMs on sampling form 7 and on the sampling frame. For example, if geographic region were an explicit stratification variable, then from the sampling frame, separate school sampling sub-frames would be constructed for each geographic region. Another example is the possible explicit stratification of small schools, discussed in Section 5.7.1. Possibly different sample designs would then be applied to each school sampling frame to select the sample of schools. In practice, the major reason for considering explicit stratification in PISA is to implement a disproportionate allocation of the school sample to the explicit strata. For example, the same number of schools could be sampled from each explicit stratum, regardless of the relative size of each stratum. The objective would be to produce equally reliable estimates for each stratum. This differs from a proportional allocation where large strata would have more sampled schools than small strata. With a proportional allocation, the sample size is often too small in small strata to obtain estimates that are sufficiently reliable.

5.6.3 Implicit Stratification

105. Implicit stratification consists essentially of sorting the school sampling frame by a set of implicit stratification variables. This sorting takes place within the explicit strata. This type of stratification is very effective for the school sample selection method proposed for PISA. It is a very simple way of ensuring a strictly proportional sample allocation of schools across all implicit strata. It can also lead to improved reliability of survey estimates, provided the implicit stratification variables being considered are correlated with PISA achievement (at the school level).

New in
PISA
2006

106. Implicit stratification can also be done at the student level by sorting student lists before they are input to KeyQuest for sampling. For countries doing the international grade option with direct student sampling, sorting the students with PISA-eligible students not in the international option grade listed first, followed by PISA-eligible students in the international option grade, followed by non-PISA eligible students in the international option grade, will ensure the expected number of sampled PISA students as well as sampled international option grade students. In other countries, it may be desirable to sort students by track or programme or language or grade, etc. to more strictly control the proportions of sampled students of each such category within selected schools. Sorting of students may have the effect of a slight reduction in sampling variance if student performance can vary according

to levels of these variables. Sorting of students before KeyQuest sampling has not been required in previous PISA cycles, nor is it required for 2006, except in the case of the international grade option. However, this idea is presented here for consideration and discussion with Westat at the time sampling form 7 is completed.

5.6.4 Stratification Guidelines

107. The selection of three to five stratification variables is usually sufficient. The following guidelines should be applied when selecting stratification variables:

- A classification variable must be available, with a valid entry, for every school on the frame in order to be considered as a stratification variable.
- The stratification variables used should be known or suspected to have relatively strong relationships with the variables of interest to be measured in the survey (school mean achievement in reading, mathematics, and/or science).
- Each school in the sampling frame should be classifiable to one and only one level of each stratification variable.
- Good measures of size should be known for the explicit strata. That is, for each explicit stratum, both the number of schools and the number of eligible students should be known.
- Defining very small strata, especially explicit strata, should be avoided because this is unlikely to improve the overall level of sampling precision. A few divisions of a continuous stratification variable usually provide all of the gains in sampling precision available from that variable.
- Preferably at least two schools should be selected in each explicit stratum to permit the computation of sampling error. This criterion usually sets an upper limit to the number of explicit strata that can be defined. In certain cases it is acceptable to select just one school from each of a few small strata, with most explicit strata providing two or more sample schools.

Please
Note

108. Stratification information must be recorded on Sampling Form 7 and sent to Westat who will forward the form to ACER.

5.7 Treatment of Small Schools

109. The information in this section is provided for information purposes only. The NPM is not required to deal with the treatment of small schools, except to discuss the issue with Westat and ACER.

110. For PISA, we identify **small schools, moderately small schools, and very small schools**. A **small school** is defined as any school whose approximate enrolment of 15-year-olds (ENR) falls below the target cluster size (TCS). The target cluster size is the number of students that are to be sampled from schools with large enrolment. The TCS is to be 35 for all schools in most countries although it can vary among explicit strata within a country if desired. All schools though, within each explicit stratum must have the same TCS. In countries that plan to sample only 150 schools, the TCS must be at least 35. A **very small school** is a small school whose ENR is less than a half the TCS (i.e. 17 or less in most countries). A **moderately small school** is a small school whose ENR is in the range of TCS/2 to TCS. Small schools in the sample can result in a reduced sample size of students for the national sample, below the desired target. Alternatively, the sample may contain many small schools, which is an administrative burden. In order to minimise these problems, the small schools in the sampling frame may require special treatment.

5.7.1 Stratification of Small Schools

111. Information on the importance of small schools will be reported on Sampling Form 6 by Westat and ACER and sent back to the NPM. This form also provides brief guidelines about the treatment of small schools. These guidelines are based on the following detailed guidelines that will be considered in the treatment of small schools in each country. Once the sampling frame is sent to Westat who will forward it to ACER, this issue will be decided upon in consultation with the NPM. The following explains the guidelines that Westat and ACER will generally follow to determine whether small school strata are needed.

112. To determine whether a stratum of small schools, a stratum of very small schools, or two strata, one of very small schools and one of moderately small schools, are needed, the following test will be applied. If none of the following conditions are true, then no small school strata are needed.

- If the percentage of students in very small schools ($ENR < TCS/2$) is 1 percent or MORE and the percentage of students in moderately small schools ($TCS/2 < ENR < TCS$) is 4 percent or MORE, then form a stratum of very small schools and a separate stratum of moderately small schools. STOP.
- If the percentage of students in very small schools ($ENR < TCS/2$) is 1 percent or MORE, a stratum for very small schools is needed, but no stratum for moderately small schools. STOP.
- If the percentage of students in very small schools ($ENR < TCS/2$) is LESS than 1 percent, and the percentage of students in moderately small schools ($TCS/2 < ENR < TCS$) is 4 percent or MORE, a combined stratum for small schools is needed which includes all very small and moderately small schools. STOP.

-
113. The small school strata can be further divided into additional explicit strata on the basis of other characteristics (e.g. region). Implicit stratification can also be used within the small school strata.

Sampling Form 6

Westat and ACER will use student enrolment data provided by the NPM on the sampling frame to determine the prevalence of small schools in their school system. Based on these enrolment data, specific sample design decisions may be required if the presence of small schools is important.

Sampling Form 6 – Example

Using a hypothetical scenario for Atlanta's PISA 2006 main study, there are less than 1 percent eligible students in very small schools (enrolment of 15-year-olds less than 18). There are also less than 1 percent of eligible students in moderately small schools, and only 1% of eligible students in small schools overall, leaving about 99 percent of eligible students in schools with at least 35 eligible students. Therefore, Atlanta has no need of extra strata for small schools.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

All of this information is reported in the example of Sampling Form 6 in Appendix A.

5.7.2 Sample Sizes for the Small School Strata

114. When small schools are explicitly stratified, great care is needed to ensure that an adequate sample of small schools is selected. At the same time, it is important to avoid selecting too many small schools, as they do not lead to many students in the assessment, which in turn means that the total school sample must be increased in order to meet the target student sample size. In particular, if there are many small schools in the sample, it will be necessary to have more than 150 sampled schools in total, to meet the PISA requirement of 4,500 assessed students.
115. The following discussion gives the approach that will be used by Westat and ACER to determine the minimum sample size of small schools (and of larger schools), that ensures that the resulting student sample is representative, yet gives some control to limit the total sample size of schools in extreme cases. Depending upon national circumstances, there may be variations to this approach, to meet the national sampling objectives while limiting the school sample size somewhat. It is important that NPMs avoid the possibility of having an

unapproved design, as a result of not handling small schools in an appropriate manner. The ensuing discussion assumes that there are strata of very small and moderately small schools identified. When there is just a single small school stratum, this is modified by ignoring the parts relating to very small schools. The formulae below also require modification when TCS is other than 35.

- a. From the complete sampling frame, find the proportions of total ENR that come from very small schools (P), moderately small schools (Q), and larger schools (those with ENR of at least TCS) (R). Thus $P+Q+R=1$.
- b. Calculate the figure L, where

$$L = 1 + (P/2).$$

Thus L is a positive number, a little more than 1.0.

- c. The minimum sample size for larger schools is equal to $150 \times R \times L$, rounded to the nearest integer. This minimum may need to be enlarged because of national considerations, such as the need to achieve minimum sample sizes for geographic regions or certain school types.
- d. Calculate the mean value of ENR for moderately small schools (MENR), and also for very small schools (VENR). Thus MENR is a number in the range of TCS/2 to TCS, and VENR is a number no greater than TCS/2.
- e. The number of schools that must be sampled from the stratum of moderately small schools is given by the formula:

$$(5250 \times Q \times L) / (\text{MENR}).$$

- f. The number of schools that must be sampled from the stratum of very small schools is given by the formula:

$$(2625 \times P \times L) / (\text{VENR}).$$

116. Consider the following example. Suppose that in Country X, the TCS is equal to 35, with 0.1 of the total enrolment of 15-year-olds in moderately small schools, and with 0.1 also in very small schools. Suppose that the average enrolment in moderately small schools is 25 students, while the average enrolment in very small schools is 10 students.

- Thus $P = 0.1$, $Q = 0.1$, $R = 0.8$, $\text{MENR} = 25$, and $\text{VENR} = 10$.
- From b, $L = 1.05$. Thus, from c, the sample size of larger schools must be at least $150 \times 0.8 \times 1.05 = 126$. That is, there must be at least 126 of the larger schools sampled.
- From e, the number of moderately small schools required is $(5250 \times 0.1 \times 1.05) / (25) = 22.05$; i.e., 22 schools.

- From f, the number of very small schools required is $(2625 \times 0.1 \times 1.05) / (10) = 27.6$; i.e., 28 schools.
- This gives a total sample size of $126 + 22 + 28 = 176$ schools, rather than just 150. The larger schools will yield an initial (i.e., before considering school and student nonresponse) sample of $126 \times 35 = 4410$ students. The moderately small schools will give an initial sample of approximately $22 \times 25 = 550$ students. The very small schools will give an initial sample size of approximately $28 \times 10 = 280$ students. Thus the total initial sample size of students is $4410 + 550 + 280 = 5240$. Recall that a sample of 150 schools, each school yielding an initial sample of 35 students, will give a total initial sample size of 5,250 students, so that our sample that has been modified to handle the small schools will yield approximately the target student sample size that we seek.

Sampling Form 7

NPMs must describe and prioritise all variables they intend to use for stratification, both explicit and implicit. They should also indicate the total number of explicit strata that will be defined, as well as the levels of each variable and the importance of each. That is, variables earlier in the list are considered more important in explaining between school variance than variables later in the list.

Sampling Form 7 – Example

In the following discussion, we show an example of stratification. Continuing with our fictitious scenario, Atlanta defined one stratification variable:

1. **State** – The six levels of this variable correspond to the six Atlanta states. The main purpose for using this stratification variable was to ensure sufficient sampling precision for the five larger states. This variable, therefore, was used as an explicit stratification variable.

In the example, urbanisation, with two levels, is shown as an implicit stratification variable within the explicit strata.

New for PISA 2006, Atlanta has indicated that they will not do any sorting of PISA students before they are entered into KeyQuest for sampling.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

All of this stratification information is reported in the example of Sampling Form 7 in Appendix A.

5.8 Assigning a Measure of Size (MOS) to each school

117. As indicated earlier, for each school on the frame an approximate enrolment of 15-year-olds in the school is needed. In some cases this will be an actual enrolment of 15-year-olds, based on administrative records for the current year. In many countries the number will

be the enrolment in the grade containing the most 15-year-olds, either for the current year or the previous year. This enrolment measure is denoted as ENR.

118. For the probability proportional to size sampling method to be used for PISA, for each school we must establish a measure of size (MOS), derived from ENR. This will be done by Westat and ACER, so what follows is for information purposes only. The MOS is given as: $MOS = MAX(ENR, TCS)$.

119. That is, the measure of size is equal to the enrolment estimate, unless the enrolment is less than the target cluster size (TCS), in which case the measure of size is set equal to the target cluster size. In most countries, $TCS = 35$, so that the MOS is equal to ENR or 35, whichever is the larger.

120. This, and the proper allocation of schools to the small school strata if they are necessary, ensures that the sample has a good representation of students from small and very small schools, but may result in a shortfall in the total student sample, unless the sample size of schools is increased to compensate.

5.9 Certainty School Stratum

121. Some countries will have schools that will be selected with certainty. This means that within their explicit sampling stratum, their MOS values were equal to or greater than the stratum sampling interval (i.e. the total cumulative stratum MOS divided by the number of schools to be sampled from the stratum (see also section C.4.1)). In such cases, the certainty selections are removed from their stratum, and the sampling interval in that stratum is recalculated for the selection of the rest of its sample. The certainty selections from all explicit strata will be put together into one explicit certainty stratum. The stratum to which each certainty selection originally belonged will be apparent from the original stratum identifier that will be available on the frame data file.

New for
PISA
2006

5.10 Strata Numbering

122. Small school strata will generally follow originally defined explicit strata in their numbering. These will in turn be followed by the certainty selection stratum. The only exception to this will be when there are separately adjudicated regions within a country. In such cases, within each separately adjudicated region, originally defined explicit strata will be followed by any small school strata for the region. These will then be followed by a stratum of certainty selections, if it exists, for the region.

5.11 Sampling Frame Statistics

123. NPMs are required to complete Sampling Forms 4, 5 and 7 to document the process of sampling frame construction. In addition, Westat and ACER will tabulate basic statistics by strata from the school sampling frame and send this back to the NPM as Sampling Form 8. For each combination of explicit and implicit strata, the number of schools and enrolled students (both ENR and MOS -- see Section 5.8) will be reported. These statistics will be very useful to finalise the sample allocation of schools to the explicit strata (Sampling Form 9) and as control totals to validate the computation of sampling weights. Sampling Form 9 (sample allocation across explicit strata) will also be completed by Westat and ACER and sent back to the NPM, as well as Sampling Form 10. Sampling Form 10 shows the information that was used by ACER to select the school sample.

Sampling Form 8

Westat and ACER will list and describe all combinations of explicit and implicit strata in Sampling Form 8. For each entry, the number of schools, eligible students, and school MOS will be tabulated. These statistics should be readily available from the school sampling frame.

Sampling Form 8 – Example

Appendix A contains a completed example of Sampling Form 8 based on the fictitious scenario for Atlanta. The example form shows the states of Atlanta as explicit strata and urbanisation as an implicit stratification variable.

The date this version of the sampling form is completed is recorded on the line beneath the NPM's name.

Sampling Form 9

Westat and ACER will describe the country's school sample allocation on Sampling Form 9. The sample allocation will show the allocation of schools to the explicit strata.

Sampling Form 9 - Example

The example depicted in Appendix A involves fictitious information for Atlanta's sample design.

The example of Sampling Form 9 shows what Atlanta's proposed stratification and allocation would be like. Explicit strata are defined based on state, and allocation is done proportionally so that the percentage of eligible students in the sample in each stratum is about the same as the percentage of eligible students in the population in each stratum. Note that the total sample size is 151 since Atlanta does not anticipate any field school exclusions, ineligibility, or nonresponse.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

Sampling Form 10

ACER will complete Sampling Form 10 to demonstrate how the school sample was selected.

Sampling Form 10 – Examples

The example of sampling form 10 is based on one of Atlanta's explicit strata from a previous example.

This scenario requires the selection of 40 schools. This example obviously consists of a single explicit stratum. In box [a] we have recorded the total MOS for that stratum (measure of size for state 2 from column 5 of the earlier Atlanta example of Sampling Form 8). In box [b] we have recorded the allocated school sample required: 40. The resulting sampling interval (see section C.4.1) is recorded in box [c] and is in fact 1490.3500. A random number is generated and recorded in box [d]. The example of Sampling Form 10 lists 40 line numbers with all the generated selection numbers needed to select the sample of 40 schools in the sampling frame for this explicit stratum.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

All such information can be found in the examples of Sampling Form 10 in Appendix A.

Sampling Form 11

NPMs must send Sampling Form 11 as a written record of their sampling frame. It should preferably be submitted as a single computer file with all the relevant information included. Once the school sample is selected by ACER, Sampling Form 11 will be sent back to the NPM with sampled and replacement schools identified and PISA IDs assigned. (See Sections 6.2 and 6.3).

Sampling Form 11 – Example

Appendix A contains an example of Sampling Form 11, again based on the fictitious scenario presented in the previous examples. Only the first eight pages are presented for the sake of illustration.

Sampled schools are determined based on the selection numbers computed in Sampling Form 10. The eight pages of the example of Sampling Form 11 included in Appendix A show the first nine sampled schools. Two replacement schools are designated for each sampled school. The PISA school IDs allow us to make the correspondence between replacement schools and sampled schools.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

All details about this example of Sampling Form 11 are found in Appendix A.

Check Point 3

NPMs must submit Sampling Forms 4, 5, 7, and 11 to Westat two months before the school sample is to be selected.

STEP 4 – THE SAMPLE OF SCHOOLS

6.1 The School Sample

124. Schools are selected once the school sampling frame has been divided into subframes for each explicit stratum, with each explicit stratum being sorted by the variables defining the implicit strata within it. Following agreement of the sampling frame received from the country (Sampling Form 11), the consortium will perform this division of the sampling frame into subframes, and sort each explicit stratum by the implicit stratification variables. NPMs should note that the school samples will be selected separately for each explicit stratum. As noted previously, each school entry in the sampling frame should contain at least:

- A school identification number;
- Values for any implicit stratification variables;
- A measure of the approximate enrolment of PISA eligible students (ENR); and
- A measure of size (MOS) (assigned by Westat and ACER).

125. If, due to the complexity of its school sample design, a country has approval from the Consortium to select its own school sample, appendix C must be read at this point.

6.2 Replacement Schools

126. NPMs must make every effort to get as many of the sampled schools to participate as possible. However, it is not always possible to obtain the participation of all sampled schools. In order to avoid the resulting sample size losses, a mechanism to identify a priori replacement schools for non-participating sampled schools is needed. A second, perhaps more important, reason for identifying replacement schools a priori is to avoid the haphazard use of alternate schools as replacements, which may actually amplify response biases. Although our approach is no guarantee of avoiding non-response biases, it will at least tend to minimise the potential for bias. The technique of identifying replacements should lead to less non-response bias than the alternative of drawing a larger sample initially, in anticipation of non-response. Each sampled school in the main survey will be assigned two replacement schools in the sampling frame by ACER. **A sampled school can never be designated as the replacement school for another sampled school.**

Please
Note

127. The following information is provided for information purposes only. Replacement schools are identified as follows. For each sampled school, the schools immediately preceding and following it in Sampling Form 11 are designated as its replacement schools. The school immediately following the sampled school will be labelled “R1” (i.e. first replacement), while the school immediately preceding the sampled school will be labelled “R2” (i.e. second replacement). The ordering of the school sampling frame by the stratification variables and by size (ENR) ensures that any sampled school’s replacements will have similar characteristics. In small countries, there may be problems encountered when trying to identify two replacement schools for each sampled school. In such cases, a replacement school could be allowed to be the potential replacement for two sampled schools (a first replacement for the preceding school, and a second replacement for the following school), but an actual replacement for only one school. It may be difficult to assign replacement schools to some very large sampled schools because the sampled schools appear very close to each other in the sampling frame. At times, it may only be possible to assign a single replacement school, and perhaps none when two consecutive schools in the sampling frame are sampled.

128. Should a sampled school happen to be the last school listed in an explicit stratum, then the two schools immediately above it are its designated replacement schools. Should a sampled school happen to be the first school listed in an explicit stratum, then the two schools immediately below it are its designated replacement schools.

6.3 School ID’s

129. In order to keep track of sampled schools and replacement schools in the PISA database, all sampled schools and replacement schools must be assigned unique identification numbers. The assigned school ID’s will be assigned by ACER to all sampled schools and replacement schools and recorded on Sampling Form 11. These IDs need to remain on all data files as they are needed at all stages of data processing and weighting.

130. School ID’s will be assigned according to these rules:

- Schools will be assigned a two-digit stratum code (corresponding to the explicit strata), and a three-digit school code.
- Sampled schools will be sequentially numbered starting from 1(one), within each explicit stratum. For example, if 150 schools are sampled from a single explicit stratum, then they are assigned IDs from 001 through 150. This numbering will continue across all implicit strata.

- First replacement schools in the main survey will be assigned the school ID of their corresponding sampled schools, but incremented by 300. For example, the first replacement school for sampled school 023 is assigned school ID 323.
- Second replacement schools in the main survey will be assigned the school ID of their corresponding sampled schools, but incremented by 600. For example, the second replacement school for sampled school 036 is assigned school ID 636, and the second replacement school for sampled school 112 is assigned school ID 712.

131. For an especially large school sample, it may not be possible to assign school IDs according to the above rules. In such a case, Westat and ACER will come up with an alternate method of assigning school IDs.

6.4 The School Tracking Form

Please
Note

132. **The School Tracking Form (Sampling Form 12) will be prepared by Westat/ACER and sent back to NPMs with the school sample. The form will help NPMs to monitor the participation status of sampled schools.** It contains basic sampling information (stratum ID, PISA School ID, ENR, and national school ID) along with empty columns to record the necessary school participation status as described in section 6.4.1. All sampled schools, as well as replacement schools assigned to a sampled school, should be recorded in Sampling Form 12, with sampled schools in the first set of columns denoted by “Original Sample”, first replacement schools in the second set of columns denoted by “First Replacement”, and second replacement schools in the third set of columns denoted by “Second Replacement”. For each set of three schools (original, first replacement, and second replacement), note in the first column of Sampling Form 12 which school participated: “O” if the originally sampled school participated, “R1” if the first replacement participated, “R2” if the second replacement participated, or “None” if none of the three schools participated.

133. NPMs should progressively update Sampling Form 12 as the participation status of schools becomes known. This form, however, will only be completed once data collection has ended since only then will NPMs know which sampled schools participated and which ones required replacement. School Tracking Forms should be submitted as soon as possible after data collection. Additionally, the school participation status information should also be recorded into KeyQuest.

6.4.1 Sampled Schools

134. NPMs must make every effort to confirm the participation of as many sampled schools as possible. This is important in order to minimise the potential for non-response

biases. As already noted, sampled schools are listed in the first set of columns of Sampling Form 12.

135. Table 3 summarizes the participation codes and their meanings.

Table 3. PARTICIPATION CODES

Code	Code Meaning
P	Participant
N1	Refusal
N2	Non-participation for other reasons
E1	No PISA-eligible students
E2	School closure
E3	School ineligible for other reasons (please specify)
EX	School exclusion

136. Once a sampled school's participation has been confirmed, a "P" should be entered in column 5 of the first row. If a sampled school is unwilling or unable to participate, then enter N1 for refusal, or N2 for non-participation due to other reasons. If the sampled school is ineligible, then record an E1 for a school with no PISA-eligible students, E2 for a school closure, and E3 for ineligible for other reasons (this latter must be explained in detail). Furthermore, enter EX for a school that belongs to one of the school-level exclusion categories recorded on Sampling Form 3 Question 2, or a school where all age-eligible students are otherwise excluded. Note that a school participation status of EX requires the school-level exclusion *type* as indicated on Sampling Form 3 Question 2 to be noted in the Remarks column on Sampling Form 12. Thus, if a sampled school is found to be a school for students with special needs and these schools have been listed as school-level exclusions in Question 2 of Sampling Form 3 on the line beginning with "1-", then the school participation status on Sampling Form 12 will be EX and "EX-1" should be recorded in the Remarks column along with the number of PISA-eligible students in the school. If a school in the sample is found to have all age-eligible students otherwise excluded and no specific school-level exclusion of this type has been previously listed on Sampling Form 3 (e.g., a school for the blind or more generally a special education school), then use EX as instructed and give the reason for exclusion in the Remarks column (as well as the number of PISA-eligible students

in the school). For this additional exclusion type, include an exclusion type number in the Remarks column, "EX-3" for example, that does not duplicate an exclusion type number already listed in Question 2 of Sampling Form 3. Use the same exclusion type number for other sampled schools that are excluded for the same reason.

6.4.2 Replacement Schools

137. After all contacts with sampled schools have been made, NPMs will need to contact replacement schools for those sampled schools that will not participate. Each sampled school that will not participate should be replaced if possible. Second replacement schools should only be used if both the corresponding first replacement school and sampled schools will not participate. **If the original sampled school is ineligible or is a type of school belonging to defined school level exclusions, do not use any replacement schools for it.** Additionally, an original school with any eligible students should not be replaced just because the number of eligible students might be smaller than expected. As for sampled schools, the participation status of approached replacement schools should be coded as described in section 6.4.1.

Sampling Form 12

NPMs must complete the Sampling Form 12 that was sent with the school sample to keep track of the participation status of all sampled schools and their replacement schools, and must send the completed form to Westat. Countries that conduct their own sampling must create as well as complete the Sampling Form 12 and send the completed form to Westat.

Sampling Form 12 – Example

Appendix A contains a partially completed Sampling Form 12, based on a fictitious scenario for Atlanta, and using fictitious identification variable values. Note the use of the "O", "R1", "R2" and "None" in the first column of the table to summarize who responded for each trio of schools.

Note the school participation status code "EX" in the last record of this example. It was not discovered until after the sample had been selected and the school contacted that this school was a school for students with special needs. An EX participation status is entered in column 5 and is further explained by the "EX-1: (4 students)" recorded in the Remarks column. This indicates clearly that this school was excluded and is the same school-level exclusion type as that defined in the first line of Question 2 in Sampling Form 3, and that there were 4 PISA-eligible students in the school. (See: Sampling Form 3 – Example in Appendix A.)

Also, the date this version of Sampling Form 12 was completed is recorded in the line below the NPM's name.

Check Point 4

A copy of Sampling Form 12 (the school tracking form) with the school participation status codes completed for each sampled original school and for each replacement school requested to participate must be submitted to Westat within one month of the end of the data collection period.

CHECKLIST OF SCHOOL SAMPLING ACTIVITIES

<i>Activity</i>	<i>School Sampling Manual Chapter/Section reference</i>	<i>Submit to IPC</i>	<i>Due Date</i>
Specify time of testing and age definition of population to be tested	Chapter 3, Section 1	Sampling Form 1 - Time of Testing and Age Definition	Submit three months before the school sample is to be selected.
Define National Desired Target Population	Chapter 3, Section 2	Sampling Form 2 - National Desired Target Population	Submit three months before the school sample is to be selected.
Define National Defined Target Population	Chapter 3, Section 3	Sampling Form 3 - National Defined Target Population	Submit three months before the school sample is to be selected.
Create and describe sampling frame	Chapter 5, Sections 1-4	Sampling Form 4 - Sampling Frame Description	Submit two months before the school sample is to be selected.
Decide on schools to be excluded from sampling frame	Chapter 3, Section 3.3.2 Chapter 5, Section 5	Sampling Form 5 - Excluded Schools	Submit two months before the school sample is to be selected.
Decide how to treat small schools	Chapter 5, Section 7	Sampling Form 6 - Treatment of Small Schools	Westat and ACER will complete and return this form to the NPM about one month before the school sample is to be selected.
Decide on explicit and implicit stratification variables	Chapter 5, Section 6	Sampling Form 7 - Stratification	Submit two months before the school sample is to be selected.
Describe population within strata	Chapter 5, Section 9	Sampling Form 8 - Population Counts by Strata	Westat and ACER will complete and return this form to the NPM when the school sample is sent to the NPM.
Allocate sample over explicit strata	Chapter 5, Section 9	Sampling Form 9 - Sample Allocation by Explicit Strata	Westat and ACER will complete and return this form to the NPM about one month before the school sample is to be selected.
Select the school sample	Chapter 5, Section 9	Sampling Form 10 - School Sample Selection	Westat and ACER will complete and return this form to the NPM when the school sample is sent to the NPM.
Identify sampled schools, replacement schools and assign PISA school IDs	Chapter 5, Section 9	Sampling Form 11 - School Sampling Frame	Submit two months before the school sample is to be selected. ACER will return this form to the NPM with sampled schools and their replacement schools identified and with PISA IDs assigned when the school sample is selected.
Create a school tracking form	Chapter 5, Section 9	Sampling Form 12 - School Tracking Form	Submit within one month of the end of the data collection period.

APPENDIX A EXAMPLES OF COMPLETED FORMS

PISA 2006 Sampling Form 1 (Example) Main Study Time of Testing and Age Definition

See Section 3.1 of School Sampling Preparation Manual.

PISA Participant: ATLANTA

National Project Manager: Mr X

Date this version of this form was completed: 7 October 2005

1. Beginning and ending dates of assessment 15 April 2006 to 15 May 2006

The main study testing period must be no more than six weeks long, and between 1 March 2006 and 31 August 2006.

2. Please confirm that the assessment start date is after the first six weeks of the academic year.

- ☒ Yes
☐ No

3. Students who will be assessed were born between 1 January 1990 and 31 December 1990
DD MM YY DD MM YY

4. As part of the PISA sampling process for your country, will you be selecting students other than those born between the dates in 3) above? (For example students in a particular grade, no matter what age.)

☐ No

☒ Yes (please describe the additional population): Students in grade 12 born in 1988

Relationship to TIMSS Population 2:

5. Is your country, or any part of your country, a participant in TIMSS 2006 for Population 2?

- ☐ No (You may skip questions 6. and 7. Done.)
☒ Yes

6. Is your institute responsible for conducting TIMSS in your country?

- ☒ Yes
☐ No

7. Are you concerned about the possible overlap between the TIMSS and PISA main study school samples in 2006? You may choose more than one reply below.

- ☐ No, not concerned because mostly the student populations are in different schools.
☐ No, not concerned because we are a big country and very few schools will be in both studies if the two samples are drawn independently.
☒ No, not concerned because we think it is acceptable for some schools to be in both studies, and we want the best possible sample for each study.
☐ Yes, it is important that as **few** schools as possible be in both studies.
☐ Yes, it is important that as **many** schools as possible be in both studies (so that the total number of schools from the two studies is as small as possible).

See Section 3.2 of *School Sampling Preparation Manual*.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 October 2005**

1. Total national population of 15-year-olds:

[a] **231645**

2.1. Total national population of 15-year-olds enrolled in educational institutions and who are in grades 7 and higher:

[b] **226658**

2.2. Total number of 15-year-olds enrolled in grades 6 and below:

None

3. If the national desired target population for your country differs from the total national population of enrolled 15-year-olds in grades 7 and higher, describe the population(s) to be omitted from the total national population of enrolled 15-year-olds in grades 7 and higher. These include distinct geographic regions of the country, or specific language groups:

None

Total enrolment omitted from the total national population of enrolled 15-year-olds in grades 7 and higher (*corresponding to the omissions listed above*):

[c] **0**

4. Total enrolment in the national desired target population:
box [b] - box [c]

[d] **226658**

5. Percentage of coverage of nationally enrolled 15-year-olds, in the national desired target population:
(box [d] / box [b]) x 100

[e] **100%**

6. Describe your data source (*i.e. year of data, provide copies of relevant tables or a webpage address*):

Statistical Yearbook of Atlanta 2004

Database of Statistical Yearbook (school year 2004/2005), www.ATL.education.gov

See Section 3.3 of School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 October 2005**

1. Total enrolment in the national desired target population:

[a] **226658**

From box [d] on Sampling Form 2

2. School-level exclusions (see Section 3.3.2):

Description of exclusions	# of schools	# of students
1 – <i>Special schools for students with special needs</i>	145	2466
2 -		
3 -		
4 -		
TOTAL	145	[b] 2466

Percentage of students not covered due to school-level exclusions:

(box [b] / box [a]) x 100

1.1 %

3. Total enrolment in national defined target population before within-school exclusions:

[c] **224192**

box [a] - box [b]

4. Anticipated within-school exclusions (students who could not be included in the PISA assessment, from schools where some students could be included):

Description of exclusions	Expected # of students
Functionally disabled students	negligible
Intellectually disabled students	negligible
Students with insufficient language experience in test language	negligible
Other	none
TOTAL.....	[d] negligible

Expected percentage of students not covered due to within-school exclusions:

(box [d] / box [a]) x 100

0 %

5. Total enrolment in national defined target population:

[e] **224192**

box [a] - (box [b] + box [d])

6. Coverage of national desired target population:

[f] **98.9%**

(box [e] / box [a]) x 100

7. Describe your data source (i.e. year of data, provide copies of relevant tables or webpage address):

2004/2005 Statistical Report of the Department of Education, Atlanta: Stationary Office, www.ATL.education.gov

See Sections 5.2 – 5.4 of *School Sampling Preparation Manual*.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2005**

1. Will a sampling frame of geographic areas be used?

- ☐ Yes Go to 2
☒ No Go to 5a

2. Specify the PSU Measure of Size to be used.

- ☐ 15-year-old student enrolment
☐ Total student enrolment
☐ Number of schools
☐ Population size
☐ Other (please describe): _____

3. Specify the school year for which enrolment data will be used for the PSU Measure of Size: _____

4. Please provide a preliminary description of the information available to construct the area frame, Please consult with Westat for support and advice in the construction and use of an area — level sampling frame.

New for
PISA
2006

5a. Define the units on the school sampling frame (i.e. whole schools, shifts, tracks, programmes, etc.)

Whole schools

5b Are the units on the school frame the same units used in previous PISA cycles? (If you are a new country to PISA in 2006, please skip this question.)

Yes

6. Specify the school estimate of enrolment (ENR) of 15-year-olds that will be used.

- ☒ 15-year-old student enrolment
☐ Applying known proportions of 15-year-olds to corresponding grade level enrolments
☐ Grade enrolment of the modal grade for 15-year-olds
☐ Total student enrolment, divided by number of grades

7. Specify the year for which enrolment data will be used for school ENR. **2004/2005**

8. Please describe any other type of frame, if any, that will be used. **None**

PISA Participant: ATLANTA

National Project Manager: **Mr X**

Use additional sheets if necessary

[illegible]

See Section 5.7 of *School Sampling Preparation Manual*.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 December 2005**

1. Enrolment in small schools:

Type of school based on enrolment	# of schools	# of students	Percentage of total enrolment
Enrolment of 15-year-old students < 18	1307	1345	[a] 0.6 %
Enrolment of 15-year-old students ≥ 18 and < 35	30	897	[b] 0.4 %
Enrolment of 15-year-old students ≥ 35	6341	221950	[c] 99 %
TOTAL	7678	224192	100 %

2. If the percentage in box [a] is 1 percent or more and the percentage in box [b] is 4 percent or more, then an explicit stratum of moderately small schools is required, AND an explicit stratum for very small schools is required. Please see section 5.7.2 to determine an appropriate school sample allocation for these strata of moderately small and very small schools..

box [a] $\geq 1\%$ and box [b] $\geq 4\%$?

☐ Yes or x No
↓

Form an explicit stratum of moderately small schools and an explicit stratum of very small schools and record this on Sampling Form 7. Done.

3. If the percentage in box [a] is 1% or more, a stratum for very small schools is needed. Please see section 5.7.2 to determine an appropriate school sample allocation for this stratum of very small schools.

box [a] $\geq 1\%$?

☐ Yes or x No
↓

Form an explicit stratum of very small schools and record this on Sampling Form 7. Done.

4. If the percentage in box [a] is less than 1 percent and the percentage in box [b] is 4 percent or more, an explicit stratum of small schools is required, but no special stratum for very small schools is required. Please see Section 5.7.2 to determine an appropriate school sample allocation for this stratum of small schools.

box [a] < 1% and box [b] $\geq 4\%$?

☐ Yes or x No
↓

Form an explicit stratum of small schools including moderately small schools and very small schools and record this on Sampling Form 7..

See Section 5.6 of School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2005**

Explicit School Stratification

1. List and describe the variables used for explicit stratification. If numeric or alpha values are used for these variables on the school frame, explain here each variable value and its meaning. Note that variables earlier in the list are thought to be more important in explaining between school variance than variables later in the list.

Explicit stratification variables		# of levels
1	6 States	6
2		
3		
4		

2. Total number of explicit strata:

6

(Note: if the number of explicit strata exceeds 99, the PISA school coding scheme will not work correctly. Consult Westat and ACER.)

Implicit School Stratification

3. List and describe the variables used for implicit stratification (If numeric or alpha values are used for these variables on the school frame, explain here each variable value and its meaning. List the variables in the order in which they will be used (i.e., sorting of schools within explicit strata, and for collapsing response cells for school-level non-response adjustment. Variables first in the list are considered more important than those lower in the list and levels of such variables will not be collapsed across unless there are extreme circumstances).

Implicit stratification variables		# of levels
1	Urban/Rural	2
2		
3		
4		

New in
PISA
2006

Implicit Student Stratification

4. List and describe the variables used for implicit student stratification, if any will be used. List the variables in the order in which they will be used (i.e., sorting of students within schools before they are input into KeyQuest for student sampling).

Implicit student stratification variables		# of levels
1	None	
2		
3		

PISA 2006 Sampling Form 8 (Example)**Population Counts by Strata**

See Section 5.9 of *School Sampling Preparation Manual*.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 January 2006**

Use additional sheets if necessary

(1) Explicit Strata	(2) Implicit Strata	(3) Schools	(4)	(5)
			Population Counts	
			Students	
			ENR	MOS
State 1	Rural	745	24 623	24 733
	Urban	1 655	54 601	54 844
	Sub-total	2 400	79 224	79 577
State 2	Rural	1 075	18 563	18 646
	Urban	1 044	40 786	40 968
	Sub-total	2 119	59 349	59 614
State 3	Rural	428	12 329	12 384
	Urban	947	27 341	27 463
	Sub-total	1 375	39 670	39 847
State 4	Rural	231	5 152	5 175
	Urban	513	11 426	11 477
	Sub-total	744	16 578	16 652
State 5	Rural	245	7 049	7 080
	Urban	541	15 629	15 699
	Sub-total	786	22 678	22 779
State 6	Rural	80	2 081	2 090
	Urban	174	4 612	4 633
	Sub-total	254	6 693	6 723
TOTAL		7 678	224 192	225 192

This form will be returned to the NPM. If a country selects its own sample, please return completed form to: *Keith Rust, at Westat*, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 9 (Example)**Sample Allocation by Explicit Strata**

See Section 5.9 and Appendix C Section C.2 of School Sampling Preparation Manual

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 December 2005**

Use additional sheets if necessary

(1) Explicit Strata	(2) Stratum ID	(3) % of eligible students in population	(4) No. of schools in population	(5) (6) (7) Sample Allocation		
				Schools	Students	% of eligible students expected in sample
State 1	01	35.34	2400	53	1855	35.10
State 2	02	26.47	2119	40	1400	26.49
State 3	03	17.69	1375	27	945	17.89
State 4	04	7.39	744	11	385	7.28
State 5	05	10.12	786	15	525	9.94
State 6	06	2.99	254	5	175	3.31
TOTAL		100.00	7678	151	5285	100.01

This form will be returned to the NPM. If a country selects its own sample, please return completed form to: *Keith Rust, at Westat*, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 10 (Example)**School Sample Selection**

See Section 5.9 and Appendix C Section C.4 of School Sampling Preparation Manual

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 January 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

S 59 614	D 40	I 1 490.3500	RN 0.6381
[a] Total Measure of Size	[b] Desired Sample Size	[c] Sampling Interval	[d] Random Number

Use additional sheets if necessary

(1) Line Numbers	(2) Selection Numbers
1	950.9923
2	2441.3423
3	3931.6923
4	5422.0423
5	6912.3923
6	8402.7423
7	9893.0923
8	11383.4423
9	12873.7923
10	14364.1423
11	15854.4923
12	17344.8423
13	18835.1923
14	20325.5423
15	21815.8923
16	23306.2423
17	24796.5923
18	26286.9423
19	27777.2923
20	29267.6423
21	30757.9923

This form will be returned to the NPM. If a country selects its own sample, please return completed form to: *Keith Rust, at Westat*, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 10 (Example)**School Sample Selection**

See Section 5.9 and Appendix C Section C.4 of School Sampling Preparation Manual

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 January, 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

S 59 614	D 40	I 1 490.3500	RN 0.6381
[a] Total Measure of Size	[b] Desired Sample Size	[c] Sampling Interval	[d] Random Number

Use additional sheets if necessary

(1) Line Numbers	(2) Selection Numbers
22	32248.3423
23	33738.6923
24	35229.0423
25	36719.3923
26	38209.7423
27	39700.0923
28	41190.4423
29	42680.7923
30	44171.1423
31	45661.4923
32	47151.8423
33	48642.1923
34	50132.5423
35	51622.8923
36	53113.2423
37	54603.5923
38	56093.9423
39	57584.2923
40	59074.6423

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed (NPM): **7 November 2006**

Date this version of this form was completed (ACER): **7 January 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
0829	Urban	110	110	110		
0552	Urban	101	101	211		
1802	Urban	98	98	309		
1288	Urban	98	98	407		
2043	Urban	95	95	502		
0974	Urban	94	94	596		
1718	Urban	94	94	690		
1807	Urban	93	93	783		
0457	Urban	93	93	876	R2	601
0244	Urban	93	93	969	S	001
1817	Urban	91	91	1 060	R1	301
1741	Urban	90	90	1 150		
1652	Urban	89	89	1 239		
0121	Urban	89	89	1 328		
0309	Urban	89	89	1 417		
0032	Urban	89	89	1 506		
0021	Urban	89	89	1 595		
0609	Urban	88	88	1 683		
0399	Urban	86	86	1 769		
0067	Urban	86	86	1 855		
0202	Urban	86	86	1 941		
0063	Urban	86	86	2 027		
1467	Urban	86	86	2 113		

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
1381	Urban	86	86	2 199		
1043	Urban	84	84	2 283		
1318	Urban	84	84	2 367	R2	602
0659	Urban	84	84	2 451	S	002
0612	Urban	83	83	2 534	R1	302
1696	Urban	82	82	2 616		
0867	Urban	82	82	2 698		
0537	Urban	81	81	2 779		
1794	Urban	80	80	2 859		
0695	Urban	80	80	2 939		
0031	Urban	80	80	3 019		
0333	Urban	79	79	3 098		
0051	Urban	79	79	3 177		
0384	Urban	79	79	3 256		
1391	Urban	79	79	3 335		
1189	Urban	79	79	3 414		
0731	Urban	78	78	3 492		
0634	Urban	78	78	3 570		
1230	Urban	77	77	3 647		
1478	Urban	76	76	3 723		
0741	Urban	76	76	3 799		
1388	Urban	76	76	3 875	R2	603
2110	Urban	76	76	3 951	S	003

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
1813	Urban	75	75	4 026	R1	303
0597	Urban	75	75	4 101		
0724	Urban	74	74	4 175		
1672	Urban	74	74	4 249		
1535	Urban	73	73	4 322		
1932	Urban	73	73	4 395		
2108	Urban	73	73	4 468		
1898	Urban	73	73	4 541		
0697	Urban	73	73	4 614		
0175	Urban	72	72	4 686		
1584	Urban	72	72	4 758		
0026	Urban	72	72	4 830		
0577	Urban	72	72	4 902		
1729	Urban	71	71	4 973		
0004	Urban	71	71	5 044		
1521	Urban	70	70	5 114		
0373	Urban	70	70	5 184		
1377	Urban	70	70	5 254		
0643	Urban	70	70	5 324		
2114	Urban	70	70	5 394	R2	604
1079	Urban	70	70	5 464	S	004
1452	Urban	70	70	5 534	R1	304
1168	Urban	70	70	5 604		

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
1591	Urban	69	69	5 673		
1309	Urban	69	69	5 742		
0661	Urban	69	69	5 811		
0017	Urban	69	69	5 880		
1174	Urban	69	69	5 949		
1983	Urban	69	69	6 018		
1766	Urban	69	69	6 087		
0236	Urban	69	69	6 156		
0412	Urban	68	68	6 224		
1032	Urban	68	68	6 292		
1676	Urban	68	68	6 360		
1030	Urban	68	68	6 428		
1085	Urban	68	68	6 496		
0158	Urban	67	67	6 563		
0834	Urban	67	67	6 630		
1135	Urban	67	67	6 697		
0154	Urban	67	67	6 764		
1279	Urban	67	67	6 831		
0504	Urban	67	67	6 898	R2	605
1858	Urban	67	67	6 965	S	005
0084	Urban	67	67	7 032	R1	305
0406	Urban	66	66	7 098		
0273	Urban	66	66	7 164		

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
0450	Urban	66	66	7 230		
1669	Urban	66	66	7 296		
0307	Urban	66	66	7 362		
1261	Urban	66	66	7 428		
0265	Urban	65	65	7 493		
1112	Urban	65	65	7 558		
1315	Urban	65	65	7 623		
0726	Urban	65	65	7 688		
1495	Urban	65	65	7 753		
0144	Urban	65	65	7 818		
1345	Urban	65	65	7 883		
1996	Urban	64	64	7 947		
1383	Urban	64	64	8 011		
1322	Urban	64	64	8 075		
0827	Urban	64	64	8 139		
1829	Urban	64	64	8 203		
0713	Urban	64	64	8 267		
1339	Urban	64	64	8 331		
1038	Urban	63	63	8 394	R2	606
0869	Urban	63	63	8 457	S	006
0740	Urban	63	63	8 520	R1	306
1784	Urban	63	63	8 583		
1157	Urban	63	63	8 646		

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
0039	Urban	63	63	8 709		
1578	Urban	63	63	8 772		
0818	Urban	62	62	8 834		
1194	Urban	62	62	8 896		
1723	Urban	62	62	8 958		
0079	Urban	62	62	9 020		
0172	Urban	62	62	9 082		
0394	Urban	62	62	9 144		
0005	Urban	62	62	9 206		
1987	Urban	61	61	9 267		
1017	Urban	61	61	9 328		
1355	Urban	61	61	9 389		
1018	Urban	61	61	9 450		
1680	Urban	61	61	9 511		
1547	Urban	60	60	9 571		
1707	Urban	60	60	9 631		
0095	Urban	60	60	9 691		
0012	Urban	60	60	9 751		
1730	Urban	60	60	9 811		
1726	Urban	60	60	9 871	R2	607
0542	Urban	60	60	9 931	S	007
0033	Urban	60	60	9 991	R1	307
0326	Urban	60	60	10 051		

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
0887	Urban	60	60	10 111		
0925	Urban	59	59	10 170		
0686	Urban	59	59	10 229		
0291	Urban	59	59	10 288		
1500	Urban	59	59	10 347		
1664	Urban	59	59	10 406		
1418	Urban	59	59	10 465		
1191	Urban	59	59	10 524		
1608	Urban	59	59	10 583		
0747	Urban	59	59	10 642		
1240	Urban	58	58	10 700		
0052	Urban	58	58	10 758		
1658	Urban	58	58	10 816		
2070	Urban	58	58	10 874		
0857	Urban	58	58	10 932		
0472	Urban	58	58	10 990		
0098	Urban	58	58	11 048		
0349	Urban	58	58	11 106		
1733	Urban	58	58	11 164		
1243	Urban	58	58	11 222		
0372	Urban	58	58	11 280		
1172	Urban	58	58	11 338	R2	608
0167	Urban	57	57	11 395	S	008

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 11 (Example)**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **7 November 2006**

Explicit Stratum: **State 2**

Stratum ID: **02**

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID
1346	Urban	57	57	11 452	R1	308
1639	Urban	57	57	11 509		
1732	Urban	57	57	11 566		
0380	Urban	57	57	11 623		
0986	Urban	57	57	11 680		
1372	Urban	56	56	11 736		
1458	Urban	56	56	11 792		
1259	Urban	56	56	11 848		
1008	Rural	56	56	11 904		
1087	Rural	60	60	11 964		
0162	Rural	64	64	12 028		
0717	Rural	65	65	12 093		
0047	Rural	66	66	12 159		
0718	Rural	70	70	12 229		
0567	Rural	93	93	12 322		
1041	Rural	150	150	12 472		
1286	Rural	165	165	12 637		
0617	Rural	170	170	12 807	R2	609
2053	Rural	170	170	12 977	S	009
0328	Rural	170	170	13 147	R1	309
1013	Rural	190	190	13 337		
1781	Rural	200	200	13 537		
0135	Rural	255	255	13 792		
...

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 12 (Example)

School Tracking Form

See Section 6.4 of School Sampling Preparation Manual

PISA Participant: **ATLANTIS**

National Project Manager: **Mr X**

Date this version of this form was completed: **15 June 2006**

Country																			
Participant	Original Sample						First Replacement						Second Replacement						
Sample	PISA_Stratum	Orig_stratum	PISA ID	School ID	ENR	Participation	PISA_Stratum	Orig_stratum	PISA ID	School ID	ENR	Participation	PISA_Stratum	Orig_Stratum	PISA ID	School ID	ENR	Participation	Remarks
Status			ID	ID		Status			ID	ID		Status			ID	ID		Status	Remarks
R1	01	01	001	W4627	28	N1	01	01	301	S4598	28	P	01	01	601	K4433	29		
O	01	01	002	A3971	27	P	01	01	302	K8324	27		01	01	602	X3409	27		
O	01	02	003	T7834	25	P	01	02	303	R6194	24		01	02	603	F2111	26		
O	01	02	004	R5610	22	P	01	02	304	G8103	21		01	02	604	D4398	23		
None	01	03	005	U8320	20	N1	01	03	305	M8215	20	E1	01	03	605	F6290	20	N1	
O	01	04	006	G6784	20	P	01	04	306	K8111	20		01	04	606	J8743	20		
O	01	04	007	A2239	20	P	01	04	307	F5550	19		01	04	607	K9911	20		
None	01	04	008	K0098	18	N2	01	04	308	V6723	18		01	04	608	M8754	19		Completed test booklets never
None	01	05	009	L9823	16	E3	01	05	309	P0098	15		01	05	609	B5600	17		
R2	01	05	010	Q3498	15	N1	01	05	310	D3495	15	N2	01	05	610	F0088	15	P	R1 failed to reply in time
O	01	05	011	W9221	14	P	01	05	311	S3337	13		01	05	611	V6021	15		
None	01	05	012	H6510	12	N1	01	05	312	A2008	11	N1	01	05			11		No schools available for 2nd replacement
O	01	05	013	J7622	10	P	01	05	313	Z3087	9		01	05	613	K0189	9		
None	01	05	014	K1784	5	EX	01	05	314	L0012	5		01	05	614	A2233	5		EX-1 [school for students with special needs] - 6 students

Please return completed form with completed to: **Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com**

APPENDIX B SAMPLING FORM TEMPLATES

See Section 3.1 of School Sampling Preparation Manual.

PISA Participant: _____

National Project Manager: _____

Date this version of this form was completed: _____

1. Beginning and ending dates of assessment _____ 2006 to _____ 2006

The main study testing period must be no more than six weeks long, and between 1 March 2006 and 31 August 2006.

2. Please confirm that the assessment start date is after the first six weeks of the academic year.

- ☐ Yes
☐ No

3. Students who will be assessed were born between _____ and _____
DD MM YY DD MM YY

4. As part of the PISA sampling process for your country, will you be selecting students other than those born between the dates in 3) above? (For example students in a particular grade, no matter what age.)

- ☐ No
☐ Yes (please describe the additional population): _____

Relationship to TIMSS Population 2:

8. Is your country, or any part of your country, a participant in TIMSS 2006 for Population 2?

- ☐ No (You may skip questions 6. and 7. Done.)
☐ Yes

9. Is your institute responsible for conducting TIMSS in your country?

- ☐ Yes
☐ No

10. Are you concerned about the possible overlap between the TIMSS and PISA main study school samples in 2006? You may choose more than one reply below.

- ☐ No, not concerned because mostly the student populations are in different schools.
☐ No, not concerned because we are a big country and very few schools will be in both studies if the two samples are drawn independently.
☐ No, not concerned because we think it is acceptable for some schools to be in both studies, and we want the best possible sample for each study.
☐ Yes, it is important that as **few** schools as possible be in both studies.
☐ Yes, it is important that as **many** schools as possible be in both studies (so that the total number of schools from the two studies is as small as possible).

See Section 3.2 of *School Sampling Preparation Manual*.

PISA Participant:

National Project Manager:

Date this version of this form was completed:

1. Total national population of 15-year-olds:

[a]

2.1. Total national population of 15-year-olds enrolled in educational institutions and who are in grades 7 and higher:

[b]

2.2. Total number of 15-year-olds enrolled in grades 6 and below:

3. If the national desired target population for your country differs from the total national population of enrolled 15-year-olds in grades 7 and higher, describe the population(s) to be omitted from the total national population of enrolled 15-year-olds in grades 7 and higher. These include distinct geographic regions of the country, or specific language groups:

Total enrolment omitted from the total national population of enrolled 15-year-olds in grades 7 and higher (*corresponding to the omissions listed above*):

[c]

4. Total enrolment in the national desired target population:
box [b] - box [c]

[d]

5. Percentage of coverage of nationally enrolled 15-year-olds, in the national desired target population:
(box [d] / box [b]) x 100

[e]

6. Describe your data source (*i.e. year of data, provide copies of relevant tables or a webpage address*):

See Section 3.3 of *School Sampling Preparation Manual*.

PISA Participant:

National Project Manager:

Date this version of this form was completed:

1. Total enrolment in the national desired target population:

From box [d] on Sampling Form 2

[a]

2. School-level exclusions (see Section 3.3.2):

Description of exclusions	# of schools	# of students
1 -		
2 -		
3 -		
4 -		
TOTAL		[b]

Percentage of students not covered due to school-level exclusions:

(box [b] / box [a]) x 100

%

3. Total enrolment in national defined target population before within-school exclusions:

box [a] - box [b]

[c]

4. Anticipated within-school exclusions (students who could not be included in the PISA assessment, from schools where some students could be included):

Description of exclusions	Expected # of students
Functionally disabled students	
Intellectually disabled students	
Students with insufficient language experience in test language	
Other	
TOTAL.....	[d]

Expected percentage of students not covered due to within-school exclusions:

(box [d] / box [a]) x 100

%

5. Total enrolment in national defined target population:

box [a] - (box [b] + box [d])

[e]

6. Coverage of national desired target population:

(box [e] / box [a]) x 100

[f]

7. Describe your data source (i.e. year of data, provide copies of relevant tables or webpage address):

See Sections 5.2 – 5.4 of *School Sampling Preparation Manual*.

PISA Participant: _____

National Project Manager: _____

Date this version of this form was completed: _____

1. Will a sampling frame of geographic areas be used?

- ☐ Yes Go to 2
☐ No Go to 5a

2. Specify the PSU Measure of Size to be used.

- ☐ 15-year-old student enrolment
☐ Total student enrolment
☐ Number of schools
☐ Population size
☐ Other (please describe): _____

3. Specify the school year for which enrolment data will be used for the PSU Measure of Size: _____

4. Please provide a preliminary description of the information available to construct the area frame, Please consult with Westat for support and advice in the construction and use of an area — level sampling frame.

New for
PISA
2006

5a. Define the units on the school sampling frame (i.e. whole schools, shifts, tracks, programmes, etc.)

5b Are the units on the school frame the same units used in previous PISA cycles? (If you are a new country to PISA in 2006, please skip this question.)

6. Specify the school estimate of enrolment (ENR) of 15-year-olds that will be used.

- ☐ 15-year-old student enrolment
☐ Applying known proportions of 15-year-olds to corresponding grade level enrolments
☐ Grade enrolment of the modal grade for 15-year-olds
☐ Total student enrolment, divided by number of grades

7. Specify the year for which enrolment data will be used for school ENR. _____

8. Please describe any other type of frame, if any, that will be used. _____

Please return completed form to: *Keith Rust, at Westat*, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

See Section 5.5 of School Sampling Preparation Manual.

PISA Participant: _____

National Project Manager:

Date this version of this form was completed: _____

Use additional sheets if necessary

[illegible]

See Section 5.7 of *School Sampling Preparation Manual*.

PISA Participant:

National Project Manager:

Date this version of this form was completed:

1. Enrolment in small schools:

Type of school based on enrolment	# of schools	# of students	Percentage of total enrolment
Enrolment of 15-year-old students < 18			[a]
Enrolment of 15-year-old students ≥ 18 and < 35			[b]
Enrolment of 15-year-old students ≥ 35			[c]
TOTAL.....			100%

2. If the percentage in box [a] is 1 percent or more and the percentage in box [b] is 4 percent or more, then an explicit stratum of moderately small schools is required, AND an explicit stratum for very small schools is required. Please see section 5.7.2 to determine an appropriate school sample allocation for these strata of moderately small and very small schools..

box [a] $\geq 1\%$ and box [b] $\geq 4\%$?

☐ Yes or ☐ No



Form an explicit stratum of moderately small schools and an explicit stratum of very small schools and record this on Sampling Form 7. Done.

3. If the percentage in box [a] is 1% or more, a stratum for very small schools is needed. Please see section 5.7.2 to determine an appropriate school sample allocation for this stratum of very small schools.

box [a] $\geq 1\%$?

☐ Yes or ☐ No



Form an explicit stratum of very small schools and record this on Sampling Form 7. Done.

4. If the percentage in box [a] is less than 1 percent and the percentage in box [b] is 4 percent or more, an explicit stratum of small schools is required, but no special stratum for very small schools is required. Please see Section 5.7.2 to determine an appropriate school sample allocation for this stratum of small schools.

box [a] < 1% and box [b] $\geq 4\%$?

☐ Yes or ☐ No



Form an explicit stratum of small schools including moderately small schools and very small schools and record this on Sampling Form 7..

See Section 5.6 of *School Sampling Preparation Manual*.

PISA Participant:

National Project Manager:

Date this version of this form was completed:

Explicit School Stratification

1. List and describe the variables used for explicit stratification. If numeric or alpha values are used for these variables on the school frame, explain here each variable value and its meaning. Note that variables earlier in the list are thought to be more important in explaining between school variance than variables later in the list.

Explicit stratification variables		# of levels
1		
2		
3		
4		

2. Total number of explicit strata:

(Note: if the number of explicit strata exceeds 99, the PISA school coding scheme will not work correctly. Consult Westat and ACER.)

Implicit School Stratification

3. List and describe the variables used for implicit stratification (If numeric or alpha values are used for these variables on the school frame, explain here each variable value and its meaning. List the variables in the order in which they will be used (i.e., sorting of schools within explicit strata, and for collapsing response cells for school-level non-response adjustment. Variables first in the list are considered more important than those lower in the list and levels of such variables will not be collapsed across unless there are extreme circumstances).

Implicit stratification variables		# of levels
1		
2		
3		
4		

New in
PISA
2006

Implicit Student Stratification

4. List and describe the variables used for implicit student stratification, if any will be used. List the variables in the order in which they will be used (i.e., sorting of students within schools before they are input into KeyQuest for student sampling).

Implicit student stratification variables		# of levels
1		
2		
3		

PISA 2006 Sampling Form 8

Population Counts by Strata

PISA 2006 Sampling Form 8

Population Counts by Strata

See Section 5.9 of School Sampling Preparation Manual.

PISA Participant: _____

National Project Manager:

Date this version of this form was completed:

Use additional sheets if necessary

[illegible]

This form will be returned to the NPM. If a country selects its own sample, please return completed form to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 9
Sample Allocation by Explicit Strata

PISA 2006 Sampling Form 9
Sample Allocation by Explicit Strata

See Section 5.9 and Appendix C Section C.2 of School Sampling Preparation Manual

PISA Participant: _____

National Project Manager:

Date this version of this form was completed:

Use additional sheets if necessary

[illegible]

This form will be returned to the NPM. If a country selects its own sample, please return completed form to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

See Section 5.9 and Appendix C Section C.4 of School Sampling Preparation Manual

PISA Participant:

National Project Manager:

Date this version of this form was completed:

Explicit Stratum:

Stratum ID:

S [a] Total Measure of Size	D [b] Desired Sample Size	I [c] Sampling Interval	RN [d] Random Number
---------------------------------------	-------------------------------------	-----------------------------------	--------------------------------

Use additional sheets if necessary

(1) Line Numbers	(2) Selection Numbers

PISA 2006 Sampling Form 11**School Sampling Frame**

See Chapter 5 and Appendix C Sections C.5 and C.6 of the School Sampling Preparation Manual.

PISA Participant:

National Project Manager:

Date this version of this form was completed:

Explicit Stratum:

Stratum ID:

Use additional sheets if necessary

(1) School List ID	(2) Implicit Stratum	(3) ENR	(4) MOS	(5) Cumulative MOS	(6) Sample Status	(7) PISA School ID

Please return form with completed columns 1-5 to: Keith Rust, Westat, 1650 Research Boulevard, Rockville, MD 20850-3195, USA Fax: 301-294-2034, Email: keithrust@Westat.com If a country selects its own sample, please return form with completed columns 1-7 to: Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com

PISA 2006 Sampling Form 12	School Tracking Form
-----------------------------------	-----------------------------

PISA 2006 Sampling Form 12	School Tracking Form
-----------------------------------	-----------------------------

See Section 6.4 of School Sampling Preparation Manual

PISA Participant:

National Project Manager:

Date this version of this form was completed:

[illegible]

Please return completed form to: *Keith Rust, at Westat, — Fax: +1-301-294-2034, E-mail: KeithRust@Westat.com*

APPENDIX C SCHOOL SAMPLING SELECTION METHODOLOGY

C.1 Overview

138. This chapter describes how to select the sample of schools using a PPS systematic sampling method, and is only relevant to those countries that have reached agreement with the Consortium to select their own school samples. It is presented as a series of operational steps leading to the selection and identification of all sampled schools.
139. The school selection probabilities are based on their measures of size (MOS). This sampling method can be applied to both school-level sampling frames as well as area-level sampling frames. With area-level sampling frames, the second level of sampling frame is very much like a school-level sampling frame, as noted previously.
140. For each sampled school, we will also identify two candidates as replacement schools. Although the replacement schools serve as possible substitutes for non-cooperating schools, NPMs should keep their use to a minimum by ensuring the highest possible participation rate for sampled schools.

C.2 School Sample Allocation Over Explicit Strata

141. If there are explicit strata, the total number of schools to be sampled in each country must be allocated among the explicit strata so that the percentage of students in the sample in each stratum is approximately the same as the percentage of students in the population in each stratum. The possibility exists that there may be some schools having a value of zero for the ENR value. Such schools are retained on the frame because they may have PISA-eligible students during the testing period. For such schools, change the ENR value to a value of one and communicate this change to Westat. The value of one should be counted towards the stratum student population percentage. NPMs must ensure that the overall sampling precision will meet PISA requirements. Based on a school sample size of 150 (or more) schools and using the information reported in columns 1, 3, and 4 of Sampling Form 8, NPMs should be able to allocate their sample of schools to the explicit strata. **This sample allocation must be reported on Sampling Form 9.**
142. NPMs are responsible for ensuring the quality execution of this school sample allocation. NPMs should consult with Westat for any assistance in the performance of this task. NPMs should pay particular attention to the following points in allocating their school sample.

- If separate estimates are required for explicit strata, then equal levels of sampling precision will usually be required for each explicit stratum. This generally means that there will be an equal number of schools sampled in each explicit stratum regardless of the size of each stratum. Countries wishing to explore this option must have their sample allocation approved by Westat.
- If different target cluster sizes are applied to the explicit strata, NPMs will need to adapt their sample allocation accordingly. Given the relative complexity in making this determination, NPMs should consult Westat for guidance.
- If one or more explicit strata of small schools are required, NPMs will need to allocate their school sample accordingly (See Section 5.7.2).
- NPMs must ensure that at least one, and preferably two, schools are allocated for selection in each explicit stratum.

Sampling Form 9

NPMs must describe their school sample allocation on Sampling Form 9. The sample allocation should show the allocation of schools to the explicit strata.

Sampling Form 9 - Example

The example depicted in Appendix A involves fictitious information for Atlanta's sample design.

The example of Sampling Form 9 shows what Atlanta's proposed stratification and allocation would be like. Explicit strata are defined based on state, and allocation is done proportionally so that the percentage of eligible students in the sample in each stratum is about the same as the percentage of eligible students in the population in each stratum.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

C.3 Sorting the Sampling Frame

143. If explicit strata are used, what follows in this section, regarding sorting of the frame, should be done within each explicit stratum. Failing to sort the sampling frame prior to sample selection will not bias or invalidate the original sample. This could increase nonresponse bias though, since replacement schools will not be as similar as possible to their originally sampled nonparticipating schools. Additionally, not sorting by the implicit stratifiers may unnecessarily increase the sampling variances of estimates obtained from the assessments, as the implicit stratification is implemented via this sort process.

144. On Sampling Form 7, NPMs defined a set of implicit stratification variables as well as a sequence from which they will be used to form implicit strata. Through this process, each school in the sampling frame will be assigned a value for each implicit stratification variable. For example, if urbanisation (rural or urban) and school type (public or private) are used, each school must be classified as either rural or urban and either as public or private.

145. The school sampling frame is then sorted by the implicit stratification variables. The schools are first sorted by the first implicit stratification variable, then by the second implicit stratification variable within the levels of the first sorting variable, and so on, until all implicit stratification variables have been exhausted. The result is a cross-classification structure of cells, where each cell represents one implicit stratum on the school sampling frame. The maximum number of implicit strata resulting from this process is equal to the product of the number of different levels of each implicit stratification variable. It is the maximum number because some cross-classifications could contain no schools at all. In our earlier example, using urbanization and school type, the cross-classification will yield a maximum of four cells ($2 \times 2 = 4$): rural-public, rural-private, urban-public, and urban-private. If private schools are only found in urban centres, then the rural-private cell will be empty, and there would only be three implicit strata.

Please
Note

146. Within each implicit stratum, the schools must be sorted by their enrolment (ENR) - NOT their MOS values. The sort order should alternate from high to low and then low to high, as indicated in the following table. For more than four implicit strata, the sort order pattern presented in the table below should simply be continued.

Implicit Stratum	Sort Order of ENR
1. Rural-Public	High to Low
2. Rural-Private	Low to High
3. Urban-Private	High to Low
4. Urban-Public	Low to High

C.4 Determining Which Schools to Sample

Please
Note

147. Sampling Form 10 is used to identify which schools are selected in the sampling frame. **There must be one Sampling Form 10 for each explicit stratum.** Be sure to indicate the explicit stratum being considered on each Sampling Form 10. This form is used to perform the basic calculations that will enable the NPMs to select the sample of schools. It will be essentially used to generate a series of selection numbers, which will be applied to the sampling frame to identify the sampled schools in Sampling Form 12.

C.4.1 Calculating the Sampling Interval

148. The PPS systematic sampling method requires the computation of a sampling interval for each explicit stratum. This calculation is performed based on information recorded in Sampling Form 10.

149. To calculate the sampling interval:

- Record the total measure of size “S” for all schools in the sampling frame for the specified explicit stratum. This number is obtained from column 5 of Sampling Form 8.
- Record this number in box [a] of Sampling Form 10.
- Record the number of schools “D” to be sampled from the specified explicit stratum. This will be the number allocated to the explicit stratum on Sampling Form 9. Record this number in box [b] of Sampling Form 10. Enter sequentially the numbers from 1 through “D” in column 1 (line numbers) of Sampling Form 10, for later use.
- Calculate the sampling interval “I” as follows: $I = S \div D$
- Record the sampling interval “I” to four decimal places in box [c] of Sampling Form 10.

C.4.2 Generating a Random Start

Please
Note

150. **Select a random number for each explicit stratum.** Random numbers are now easily generated by computer software, such as Excel, SAS or SPSS. NPMs may also rely on a table of random numbers found in some sampling books. NPMs who will generate these random numbers by computer software should be sure to use a “uniform random number generator” that generates numbers distributed uniformly between 0 and 1.

151. The generated random number “RN” should be recorded in box [d] of Sampling Form 10. The selected random number should be a number between 0 and 1 and should be recorded to four decimal places.

C.4.3 Calculating the Selection Numbers

152. Selection numbers must be calculated and recorded in column 2 of Sampling Form 10 for each line number specified, i.e., 1 through “D”.

- To obtain the first selection number, multiply the sampling interval “I” by the random number “RN”. Record this first selection number to four decimal places in column 2 of the first line, corresponding to line number 1, on Sampling Form 10. This first selection number will be used to identify the first sampled school in the specified explicit stratum.
- We obtain the second selection number by simply adding the sampling interval “I” to the first selection number. The second selection number will be used to identify the second sampled school. Record this number to four decimal places on the second line of column 2.
- Continue adding the sampling interval “I” to the previous selection number to obtain the next selection number. This is done until all specified line numbers (1 through “D”) have been assigned a selection number.

153. To summarize the calculation of selection numbers, the first selection number in an explicit stratum is $RN \times I$. The second selection number is $(RN \times I) + I$. The third selection number is $(RN \times I) + I + I$, and so on.

154. Selection numbers need to be generated independently for each explicit stratum. Be sure to select a new random number for each explicit stratum. For each explicit stratum, the first line number in column 1 of Sampling Form 10 should always be 1 (one).

Sampling Form 10

NPMs must complete Sampling Form 10 to demonstrate how the school sample is selected. Separate sampling forms are required for each explicit stratum.

Sampling Form 10 – Examples

The example is based on a fictitious scenario for one of Atlanta's explicit strata from a previous example.

This scenario requires the selection of 40 schools. This example obviously consists of a single explicit stratum. In box [a] we have recorded the total MOS for that stratum (measure of size for state 2 from column 5 of the earlier Atlanta example of Sampling Form 8). In box [b] we have recorded the allocated school sample required: 40. The resulting sampling interval is recorded in box [c] and is in fact 1490.3500. A random number is generated and recorded in box [d]. The example of Sampling Form 10 lists 40 line numbers with all the generated selection numbers needed to select the sample of 40 schools in the sampling frame for this explicit stratum.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

All such information can be found in the examples of Sampling Form 10 in Appendix A.

C.5 Identifying the Sampled Schools

155. Based on the selection numbers recorded in Sampling Form 10, we are now able to identify the sampled schools in the school sampling frame. **Since selection numbers were selected separately for each explicit stratum, it is important to match the correct selection numbers to the corresponding explicit strata in each sampling frame.**

156. Our first task, however is to compile a cumulative measure of size in the school sampling frame. This cumulative MOS will determine which schools are sampled. The cumulative MOS is tabulated within explicit strata. This means the cumulative MOS must be reset when moving from one explicit stratum to the next.

Please
Note

157. **All of the relevant school sampling frame information must be recorded on Sampling Form 11.** It is expected that the contents of Sampling Form 11 will actually exist in the form of a computer file, perhaps Excel, dBASE, SAS or SPSS. All of these formats are acceptable and NPMs are encouraged to send such files to Westat as a record of their sample selection activities, rather than paper versions of Sampling Form 11.

158. Sampled schools are identified as follows, once all schools within the explicit stratum have had columns 1-5 completed on Sampling Form 11:

159. Let "Z" denote the first selection number for a particular explicit stratum. We must find the first school in the sampling frame whose cumulative MOS equals or exceeds "Z". This will be the first sampled school. In other words, if C_s is the cumulative MOS of a particular school in the sampling frame and $C_{(s-1)}$ is the cumulative MOS of the school immediately preceding it, then the school in question is selected if:

- C_s is greater than or equal to Z , and
- $C_{(s-1)}$ is strictly less than Z .

160. Apply the above rule to all selection numbers for a given explicit stratum and enter “S” for each sampled school in column 6 of Sampling Form 11.

161. **Note:** If NPMs observe schools in their sampling frame whose MOS is greater than the computed sampling interval, they will encounter some sampling difficulties since such schools could be selected more than once. If this occurs, NPMs should notify Westat for advice on how to resolve this issue.

C.6 Identifying Replacement Schools

162. NPMs must make every effort to get as many of the sampled schools to participate as possible. However, it is not always possible to obtain the participation of all sampled schools. In order to avoid the resulting sample size losses, a mechanism to identify a priori replacement schools for non-participating sampled schools is needed. A second, perhaps more important, reason for identifying replacement schools a priori is to avoid the haphazard use of alternate schools as replacements, which may actually amplify response biases. Although our approach is no guarantee of avoiding non-response biases, it will at least tend to minimize the potential for bias. The technique of identifying replacements should lead to less nonresponse bias than the alternative of drawing a larger sample initially, in anticipation of nonresponse. Each sampled school in the main survey will be assigned two replacement schools in the sampling frame. **A sampled school can never be designated as the replacement school for another sampled school.**

163. Replacement schools are identified as follows. For each sampled school, identified by an S in column 6 of Sampling Form 11, the schools immediately preceding and following it in Sampling Form 11 are designated as its replacement schools. The school immediately following the sampled school should be labelled “R1” (i.e. first replacement), while the school immediately preceding the sampled school should be labelled “R2” (i.e. second replacement), in column 6 of Sampling Form 11. The use of implicit stratification variables, and the subsequent ordering of the school sampling frame by size (ENR), ensures that any sampled school’s replacements will have similar characteristics. In small countries, there may be problems encountered when trying to identify two replacement schools for each sampled school. In such cases, a replacement school could be allowed to be the potential replacement for two sampled schools (a first replacement for the preceding school, and a second

replacement for the following school), but an actual replacement for only one school. NPMs may find it difficult to assign replacement schools to some very large sampled schools because the sampled schools appear very close to each other in the sampling frame. At times, NPMs may only be able to assign a single replacement school, and perhaps none when two consecutive schools in the sampling frame are sampled.

164. Should a sampled school happen to be the last school listed in an explicit stratum, then the two schools immediately above it are its designated replacement schools. Should a sampled school happen to be the first school listed in an explicit stratum, then the two schools immediately below it are its designated replacement schools.

C.7 Assigning School ID's

165. In order to keep track of sampled schools and replacement schools in the PISA database, all sampled schools and replacement schools must be assigned unique identification numbers. **The assigned school ID's must be assigned to all sampled schools and replacement schools and recorded in column 7 on Sampling Form 11.**

Please
Note

166. School ID's must be assigned according to these rules:

- Schools may be assigned a two-digit stratum code (corresponding to the explicit strata), and a three-digit school code.
- Sampled schools are sequentially numbered starting from 1(one), within each explicit stratum. For example, if 150 schools are sampled from a single explicit stratum, then they are assigned IDs from 001 through 150. This numbering should continue across all implicit strata.
- First replacement schools in the main survey are assigned the school ID of their corresponding sampled schools, but incremented by 300. For example, the first replacement school for sampled school 023 is assigned school ID 323.
- Second replacement schools in the main survey are assigned the school ID of their corresponding sampled schools, but incremented by 600. For example, the second replacement school for sampled school 136 is assigned school ID 736.

167. NPMs who will select especially large school samples may not be able to assign school IDs according to the above rules. They will need to consult with Westat and ACER to come up with an alternate method of assigning school IDs.

Sampling Form 11

NPMs must complete Sampling Form 11 as a written record of their sampling frame and school sample selection. Sampling Form 11 should preferably be submitted as a computer file with all the relevant information included. There must be as many Sampling 11 Forms as there are explicit strata.

NPMs should be particularly careful when assigning PISA school IDs since this will be the only method of recognizing schools that are sampled. It is also how we will identify replacement schools from sampled schools.

Sampling Form 11 – Example

Appendix A contains an example of Sampling Form 11, again based on the fictitious scenario presented in the previous examples. Only the first eight pages are presented for the sake of illustration.

Sampled schools are determined based on the selection numbers computed in Sampling Form 10. The eight pages of the example of Sampling Form 11 included in Appendix A show the first nine sampled schools. Two replacement schools are designated for each sampled school. The PISA school IDs allow us to make the correspondence between replacement schools and sampled schools.

The date this version of the sampling form was completed is recorded on the line beneath the NPM's name.

All details about this example of Sampling Form 11 are found in Appendix A.

Check Point

Sampling Form 11 must be submitted to Westat three months before the before the start of data collection for those countries selecting their own samples.

168. Return to section 6.4.

APPENDIX D OVERLAP CONTROL – PISA AND TIMSS

D.1 Overview

169. This chapter describes procedures for controlling overlap of schools sampled in PISA and schools sampled in TIMSS. Because the main study for PISA 2006 and the 2007 Trends in International Mathematics and Science Study (TIMSS) would occur at approximately the same time in southern hemisphere countries, an overlap control procedure can be used for countries who wish for there to be a minimum (or a maximum) of the same schools to be sampled for each study. This procedure can only be done if the same school identifiers are used on the TIMSS and PISA school frames and if the schools used on each frame are the same.
170. In 2003, the TIMSS samples were usually selected before the PISA samples and we expect the same for 2006. Thus, for countries requesting overlap control, the TIMSS International Study Center will supply us with their school frames, with the school IDs, the school probability of selection, and an indicator showing which schools have been sampled for TIMSS. This information will be used to make adjustments to PISA school probabilities as discussed in D.2. In the case when PISA selects the school samples first, schools are sampled as usual unless there are any PISA school probabilities of selection greater than 0.5 (see discussion below).
171. Both TIMSS and PISA sample selections can generally avoid overlap of schools if any schools which would have been selected with high probability for either study have their probabilities capped at 0.5. Such an action makes each study's sample slightly less than optimal, but this is deemed acceptable when weighed against the possibility of low response rates due to school burden. Each study's NPM needs to decide if this is the path they wish to adopt. If they decide against this capping of probabilities, then it is possible for some large schools to be in both the TIMSS and PISA samples.

D.2 The Overlap Control Method

172. To control overlap, the sample selection of schools for PISA adopts a modification of the approach due to Keyfitz (1951), based on Bayes Theorem.
173. Suppose that $PROBT$ is the TIMSS probability of selection, and $PROBP$ is the usual PISA probability of selection (where $PROBP = MOS / \text{stratum sampling interval}$), then a conditional probability of selection into PISA, $CPROB$ is determined, based upon whether

there is a desire to minimise or maximise the overlap between the TIMSS and PISA samples. If the desire is to minimise the overlap then CPROB is defined as follows:

$$CPROB = \begin{cases} \max \left[0, \left(\frac{PROBT + PROBP - 1}{PROBT} \right) \right] & \text{if the school was a TIMSS school} \\ \min \left[1, \frac{PROBP}{(1 - PROBT)} \right] & \text{if the school was not a TIMSS school} \\ PROBP & \text{if the school was not a TIMSS eligible school} \end{cases} \quad (1)$$

174. If the desire is to maximise the overlap then CPROB is defined as follows:

$$CPROB = \begin{cases} \min \left[1, \left(\frac{PROBP}{PROBT} \right) \right] & \text{if the school was a TIMSS school} \\ \max \left[0, \frac{(PROBP - PROBT)}{(1 - PROBT)} \right] & \text{if the school was not a TIMSS school} \\ PROBP & \text{if the school was not a TIMSS eligible school} \end{cases} \quad (2)$$

175. Then a conditional MOS variable is created to coincide with these conditional probabilities as follows:

$$CMOS = CPROB \times \text{stratum sampling interval (rounded to 4 decimal places)}.$$

176. The PISA school sample is then selected using the line numbers created as usual (see C.4.2), but applied to the cumulated CMOS values (as opposed to the cumulated MOS values) (see C.5). Note that it is possible that the resulting PISA sample size could be a bit lower or higher than the originally assigned sample size, but this is deemed acceptable.